

المعالجة الرقمية للمشاريع العربية المعجمية المُحوسبة

عرض وقراءة لجهود عربية مُنجزَة

Digital processing of computerized Arabic lexical projects
Presentation and reading of accomplished Arab effortsد.إيمان شاشة¹، د.آمنة مناع²، د.تاويريرت حسام الدين³¹ وحدة البحث اللساني وقضايا اللغة العربية بالجزائر. ورقة (الجزائر)،

imane.chacha90@gmail.com

² وحدة البحث اللساني وقضايا اللغة العربية بالجزائر. ورقة (الجزائر)، menaa30@gmail.com³ وحدة البحث اللساني وقضايا اللغة العربية بالجزائر. ورقة (الجزائر)، dr.taouririt@gmail.com

تاريخ الاستلام: 2021/4/28 تاريخ القبول: 2021/6/21 تاريخ النشر: 2022/5/19

ملخص: نسعى من خلال هذه المداخلة إلى عرض أهم التجارب العربية الرائدة في مجال صناعة المعاجم الإلكترونية والمدونات، وما يرتبط بذلك من معالجة لغوية آلية لنصوص المدونة ضمن ما يسمى بالمعالجة الآلية للغات الطبيعية (NPL). وذلك من خلال التطرق لبعض من أهم التجارب في التعامل مع المدونات اللغوية الإلكترونية وبنائها وتطويرها للاستفادة منها في مرحلة لاحقة في بناء معاجم إلكترونية وبرمجيات لغوية مختلفة. على غرار تجربة معجم الدوحة التاريخي للمركز العربي للأبحاث ودراسة السياسات؛ المعجم الإلكتروني التفاعلي والمدونة اللغوية العربية لمدينة الملك عبد العزيز للعلوم والتقنية؛ المدونة اللغوية العربية العالمية لمكتبة الإسكندرية. **الكلمات المفتاحية:** المعالجة الآلية للغة؛ المعجم الإلكتروني؛ اللسانيات الحاسوبية؛ المعجم التاريخي؛ المدونة اللغوية العربية العالمية.

Abstract: We seek through this intervention to present the most important pioneering Arab experiences in the field of electronic dictionaries and corpus industry, and what is related to the automatic processing of the corpus texts within the so-called natural language processing. by addressing some of the most important experiences in dealing with electronic linguistic corpus, building and developing it to be used at a later stage in building electronic dictionaries and various linguistic software.

Similar to the experience of the Doha Historical Dictionary of the Arab Center for Research and Policy Studies; The Interactive Arabic Dictionary and the Arabic linguistic corpus for the King Abdulaziz City for Science and Technology; The International Corpus of Arabic for the Library of Alexandria.

Keywords: Natural Language Processing; Electronic Dictionary; Computational Linguistics; Historical Dictionary; International Corpus of Arabic.

1. مقدمة

اللسانيات الحاسوبية هي العلم الذي يُعنى بتوجيه الأنظمة الحاسوبية إلى فهم اللغات الطبيعية ومحاكاة الذكاء البشري، واستثمار التقانة الحديثة، وخاصة الحاسوب بما يحتويه من إمكانات رياضية خارقة، وسعة تخزينية هائلة في خدمة اللغة، وليست الدراسات اللغوية العربية بمنأى عن هذه الثورة التكنولوجية، وجعلها تواكب الدراسات اللسانية الغربية من حيث الرقمة وأرشفة التراث، هذا من خلال استخدام التكنولوجيا في التعامل مع النصوص نهوضاً بالبحث اللغوي وتطويره. وما لللسانيات الحاسوبية من أثر براغماتي كبير في تعليم اللغة العربية لأبنائها وغيرهم. فكانت مساحة استفادة المستوى المعجمي من الحاسوب واسعة جداً، الأمر الذي ساعد في ظهور ما يسمى بالمعجم الحاسوبية أو المعاجم الآلية، بل أخذ بالظهور على أنه علم مستقل، أو فرع من فروع علم اللغة الحاسوبي يطلق عليه علم المعجم الحاسوبي، واجتهد الحاسوبيون واللسانيون العرب وغيرهم في ابتكار مشاريع جادة تحاكي المعاجم الإلكترونية والمدونات ذات المعالجة الآلية للمعطيات اللغوية بمختلف مستوياتها. الأمر الذي دفعنا إلى البحث في الموضوع التالي: " المعالجة الرقمية للمشاريع المعجمية العربية - قراءة في مشاريع مُنجزّة -" في محاولة منا إلى تسليط الضوء على أهم البرمجيات الحاسوبية في الصناعة العربية للمعجم الإلكتروني، والمشاريع العربية الرائدة سواء المنجزّة منها أو التي في طور الإنجاز، نذكر على سبيل المثال لا الحصر: معجم الدوحة التاريخي، والمدونة اللغوية العربية العالمية، وغيرها من المشاريع المتميزة في رقمنة اللغة العربية وحوسبتها على نطاق واسع.

2. مصطلحات الدراسة

وقبل التطرق لعدد من التجارب الرائدة في هذا المجال، وجب علينا التعريف بعدد من المصطلحات التي يكثر استعمالها ويتردد ذكرها في هذا الباب؛ فمن ذلك نذكر:

1.2. المعالجة الآلية للغة

نظراً للتطور التكنولوجي الكبير الذي أحدثه استخدام الحاسوب في مختلف العلوم والتخصصات وانعكاسه على جميع نواحي ومجالات الحياة؛ صارت عملية تسخير علوم الحاسوب في خدمة اللغة العربية أمراً ضرورياً ولا سبيل إلى الاستغناء

عنه، فبالإضافة إلى ما توفره الحوسبة من تسهيل وتحسين للأداء وكذا ربح للوقت، فإنها كذلك وسيلة لتطوير اللغة العربية وجعلها لغة عالمية، من خلال برمجيات متنوعة تختص بمعالجة اللغة العربية وفق عدة مستويات لغوية ضمن ما يسمى بالمعالجة الآلية للغة (تاويريت، 2020، صفحة 304).

ويمكن تعريف المعالجة الآلية للغة بأنها: ((فرع من فروع الذكاء الاصطناعي واللسانيات، تهتم بتحليل، توليد وفهم لغات الإنسان. يتم التوليد انطلاقاً من قواعد بيانات تحوي معلومات مهيكلة بطريقة حاسوبية للحصول على جمل مفهومة من الإنسان. وكذا فهم لغة الإنسان الطبيعية وتحويلها إلى تمثيلات مجردة قابلة للاستغلال من طرف برامج الحاسوب)) (بسو، 2016، صفحة 5).

وعليه فإن المعالجة الآلية للغة تتيح إمكانية الانتقال بالمحتوى اللغوي من شكل يحاكي المادة الخام إلى نتاج لغوي يؤدي غاية ويحمل دلالة.

والاستفادة من ذلك في حقل الترجمة والتعليمات إضافة إلى تعزيز المحتوى الرقمي اللغوي العربي لينافس اللغات العالمية، ولعل من أهم التطبيقات الحديثة في تحقيق ذلك، المدونات اللغوية الرقمية وقواعد البيانات التي تحوي عشرات الآلاف من المدخل والمفردات العربية.

2.2. المدونة اللغوية

المدونة اللغوية ((هي نصوص إلكترونية تجمع لغرض معين بناء على معايير خارجيه. والمقصود بالمعايير الخارجية أي المعايير التي ليس لمحتوى النص علاقة بها، مثل أوعية النشر أو تاريخ النشر)) (صالح، 2021).

ويتم تخزين النصوص وحفظها رقمياً، في شكل قاعدة بيانات لتسهيل الرجوع إليها ومن أجل تحليلها ومعالجتها في وقت لاحق.

وغالباً ما تكون مثل هذه المدونات بمثابة البيئة المناسبة والأرض الخصبة لإقامة مشاريع ضخمة وإنجازات حضارية يخلدها التاريخ كمثال مشروع: المعجم التاريخي والمعجم الحاسوبي التفاعلي للغة العربية.

3.2. المعجم الإلكتروني

وهو عبارة عن نسخة حاسوبية معدلة من النسخة الورقية، فهو يتكون من عدد كبير من المداخل التي يحتوي كل واحد منها على المعلومات التي يمكن تجميعها، بحيث تختلف هذه المعلومات من معجم لآخر حسب الأهداف التي بني لأجلها وأصناف المستخدمين المستهدفين (بن حامدو، 2011، صفحة 35).

ويرى صابر الجمعاوي بأن المعجم الإلكتروني يعد شكلاً حادثاً في مجال الصناعة المعجمية، وذلك يرجع لكونه مظهراً من مظاهر توظيف الحاسوب في خدمة الظاهرة اللغوية، ويعرفه بذلك على أنه ((مخزون من المفردات اللغوية المرفقة بمعلومات عنها، ككيفية النطق بها وأصلها واستعمالاتها...، ويقوم الحاسوب بإدارته وفق برامج محددة سلفاً. ومن خصائصه أنه يمكن ولوجه واستعماله وتغييره بالحذف أو الإضافة أو غيرها، ويتميز بسهولة الاستعمال والسرعة والاسترجاع)) (الجمعاوي، 2013، صفحة 15).

3. التجارب العربية الرائدة في مجال المعالجة الرقمية للمحتوى اللغوي

ونعرض فيما يأتي بعضاً من أهم التجارب العربية الرائدة في مجال الصناعة اللغوية ورقمنة المحتوى اللغوي ومعالجته، ويتعلق الأمر بكل من مشروع: معجم الدوحة التاريخي للمركز العربي للأبحاث ودراسة السياسات؛ المعجم الإلكتروني التفاعلي والمدونة اللغوية العربية لمدينة الملك عبد العزيز للعلوم والتقنية؛ المدونة اللغوية العربية العالمية لمكتبة الإسكندرية.

1.3. معجم الدوحة التاريخي

1.1.3. تعريف المعجم التاريخي

هو معجم يرصد دلالة ألفاظ اللغة العربية في حياتها، فيتضمن (ذاكرة) كل لفظ من ألفاظ اللغة العربية، ويسجل حسب المتاح من المعلومات، تاريخ ظهوره بدلالته الأولى، وتاريخ تحولاته الدلالية، ومكان ظهوره، ومستعمليه في تطوراتها ما أمكن ذلك، مع توثيق تلك (الذاكرة) بالنصوص التي تشهد على صحة المعلومات الواردة فيها (مجموعة من المؤلفين، 2014، صفحة 22).

2.1.3. تعريف مؤسسة معجم الدوحة التاريخي

يتضمن موقع المعجم على الشابكة تعريفا موجزا له؛ حيث تم تعريف معجم الدوحة التاريخي للغة العربية على أنه ((مؤسسة لغوية علمية ذات شخصية اعتبارية، تابعة للمركز العربي للأبحاث ودراسة السياسات الدوحة، قطر. وعنوانها: المركز العربي للأبحاث ودراسة السياسات، ص.ب 200592، شارع الطرف، المنطقة 70، منطقة وادي البنات- الدوحة، قطر)) (مؤسسة معجم الدوحة التاريخي، 2021). تسعى المؤسسة من وراء العمل على هذا المعجم إلى إنجاز مدونة لغوية عربية شاملة يستفاد منها في إعداد هذا المعجم ومعاجم فرعية أخرى، إضافة لبناء برمجيات وتطبيقات مختلفة للمعالجة الآلية للغة، وتعتمد المؤسسة في عمل ذلك على طاقم من الخبراء اللغويين والحاسوبيين المؤهلين لتنفيذ المعجم بإشراف عدد من الخبراء والكفاءات في العمل المعجمي.

3.1.3. المدونة اللغوية للمعجم

تشمل مدونة المعجم جميع النصوص اللغوية المكتوبة باللغة العربية التي تعبر عن المراحل الزمنية المختلفة التي مرت بها اللغة العربية، وهي بمثابة المادة الخام التي تستقى منها مداخل المعجم ومختلف شواهدده. وتعد أولى المراحل التي تمر لها عملية بناء المعجم وتتأسس عليها، وقد احتوت مدونة معجم الدوحة التاريخي على كم كبير النصوص ويرجع ذلك لضرورة تعبير نصوص المدونة للمراحل المختلفة من تاريخ العربية بما يعكس التنوع والتغير اللغوي والتطور الدلالي لألفاظ العربية على مر الحقب الزمنية لتاريخها ((في بيئاتها ومراكزها الثقافيّة والحضاريّة التي شهِدَت مراحلَ نُموّها وتَطوُّرها)) (مؤسسة معجم الدوحة التاريخي، عن المدونة اللغوية، 2021).

4.1.3. البحث في المحتوى اللغوي للمعجم

يوفر المعجم ثلاث خيارات للبحث في المحتوى: البحث في المدونة اللغوية - البحث في مصادر المعجم . البحث المعجم.

1.4.1.3. البحث في المدونة:

يتيح المعجم البحث في نصوص المدونة المختلفة وتصفح محتواها، ويتم عرض نتائج البحث في شكل ثلاث خانات: الأولى تحتوي على النصوص التي تتضمن المدخل المراد البحث والثانية نسبة النص الذي تضمن المدخل والثالثة تتضمن الحقبة الزمنية التي يندرج ضمنها النص.

ويتم ترتيب نتائج البحث وفق تسلسلها الزمني من الأقدم إلى الأحدث، مع إتاحة إمكانية عرض النص الكامل للجملة التي تضمنت المدخل المراد البحث عنه. مما نلاحظه على محرك البحث في المدونة اللغوية أنه لا يحتوي خيارات بحث تفصيلية مساعدة لتصنيف نتائج البحث، إذ قد يتم عرض آلاف النتائج في عملية البحث الواحدة دون إمكانية تحديد لهذه النتائج، ما قد يسبب عناء للقارئ.

عدا ذلك؛ فإن المدونة تعد إنجازا لغويا ذا قيمة علمية عالية، سيمهد الطريق لإنجاز مشاريع لغوية عديدة بما يساهم في تطوير اللغة العربية وخدمة متكلميها ومتعلميها.

2.4.1.3. البحث في مصادر المعجم (البيبليوغرافيا):

يتيح المعجم البحث في قائمة المصادر التي تم استقاء مادة المدونة اللغوية منها، وتقدم صفحة البحث خيارات تفصيلية للبحث، بإتاحة إمكانية البحث عن عنوان كتاب أو مؤلف مع إمكانية تحديد الفترة الزمنية.

وتتضمن قائمة مصادر المعجم والمدونة جميع بيانات المؤلف والكتاب والمحقق والناشر وتاريخ النشر مع ذكر سنة وفاة المؤلف أو سنة تأليف الكتاب وعصره. ويتم ترتيبها في نتائج البحث ترتيبا زمنيا من الأقدم إلى الأحدث (مؤسسة معجم الدوحة التاريخي، عن مصادر المعجم، 2021).

كما تتيح هذه الصفحة أيضا البحث في مصادر النقوش ومصادر النظائر السامية المعتمدة التي تم الرجوع إليها في تأليف المعجم ومدونته اللغوية.

3.4.1.3. البحث في المعجم:

يتيح المعجم البحث في محتوى نصوصه المعجمية وفق خيارات بحث متنوعة، فنجد أن المعجم يتيح في صفحة البحث الرئيسية عند الضغط على خاصية البحث المتقدم إمكانية بحسب التخصص أو الوسم بتحديد نوع المدخل، كما يوفر قائمة

لجذور المعجم مع توفير خيار تحديد البحث في المصطلحات أو الجذور التي لها نظائر سامية ونقوش وفوق ذلك كله فالمعجم يتيح إمكانية تحديد الفترة الزمنية المعنية بالبحث، ما يوفر تجربة تصفح وبحث متميزة تقدم للمتصفح أحسن خدمة.

أما بالنسبة لنتائج البحث فترتب زمنياً، من الأقدم إلى الأحدث؛ فيتم تقديم المعنى المقابل للفظ ليتبعه الشاهد أو دليل الاستعمال مرفقاً بصاحبه والتوثيق الكامل للمرجع الذي ورد فيه، مع الإشارة إلى نوع المدخل والفترة الزمنية التي ينسب إليها، ثم يلي ذلك عرض بقية المداخل المرتبطة بالمدخل.

وتتضمن الصفحة الرئيسية للمعجم عدداً من الإحصائيات حول تواتر استخدام ألفاظ المعجم وأحدث الكلمات التي يتم إدراجها إضافة إلى تحديد للكلمات الأكثر بحثاً على موقع المعجم، والمعجم في هذا يقتدي أثر المؤسسات المعجمية العالمية في المجال، بتحديد احتياجات القراء وما يبحثون عنه لتحسين الخدمة وتطوير المعاجم والاستفادة من ذلك في مجال التسويق اللغوي.

5.1.3. المعالجة الآلية لنصوص المدونة اللغوية

1.5.1.3. تهيئة نصوص المدونة للمعالجة اللغوية الرقمية

حرصت الهيئة التنفيذية لمعجم الدوحة التاريخي على العمل بإستراتيجية الدقة الإحاطة في صياغة المدونة اللغوية للمعجم وذلك من خلال محاولة استيعاب نسبة معتبرة من النصوص الممثلة لكل حقبة زمنية، شرط التثبت من صحة ودقة النصوص والتحقق من نسبتها لأصحابها واستبعاد ما وقع فيه الشك.

. ولتحقيق هذا الهدف تم تحديد القائمة المرجعية لمصادر نصوص المدونة (البيبليوغرافيا) بما يوافق الأهداف المسطرة لبناء المدونة التي تعد الأساس لبناء المعجم التاريخي كمرحلة أولى.

. بعد تحديد قائمة المصادر الممثلة لمراحل عمر اللغة العربية، تم تحديد النصوص النموذجية التي تمثل استخدام اللغة العربية في العصور التي تنتمي إليها وانتقاؤها بعناية ووفق شروط صارمة واستثنائية في بعض الأحيان باعتماد معياري الشمول والدقة، لأن هذه النصوص هي ما سيكون المادة المكونة للمدونة اللغوية التي سيتم انتقاء مداخل وشواهد المعجم التاريخي منها.

. تحددت المرحلة الموالية في تحرير نصوص المدونة رقمياً بشكل يجعلها قابلة للقراءة الحاسوبية والتعديل وجدولتها لتمكين القراءة الإحصائية تمهيداً لمرحلة المعالجة الرقمية. وتم في هذه المرحلة الاستفادة مما توافر من مصادر إلكترونية متوفرة على الشبكة ربحاً للوقت والجهد وذلك بعد تهيئتها ومراجعتها والتأكد من مطابقتها للنسخ الورقية ثم ضبطها وفق شكل موحد بعد حذف الرموز والحواشي والشروح والحروف الغير عربية والإبقاء على النص النموذج للحقبة الزمنية تمهيداً للمعالجة اللغوية والإحصائية.

2.5.1.3. المعالجة اللغوية الرقمية لنصوص المدونة

إن المعالجة اللغوية لنصوص المدونات اللغوية تقتض مسبقاً وجود عدد من التطبيقات والبرمجيات المتعلقة بالمعالجة والتحليل اللغوي والتعرف الآلي على اللغة والتدقيق؛ أهم هذه البرمجيات: المدقق الإملائي، المحلل الصرفي، المُشكّل الآلي، نظام الاشتقاق والتصريف، نظام البحث والعرض.

وبالرغم من الجهود الفردية والجماعية في هذا الباب من أجل تطوير برمجيات المعالجة الآلية للغة العربية، إلا أنها لا تزال دون المطلوب بالنظر إلى ما هو مطلوب وما ينبغي أن يكون لتيسير المعالجة ودقتها للاستغناء عن المراجعة اليدوية.

وأمام هذا الموقف، ونظراً لحاجة المعجم التاريخي ((لمحلّ صرفي ناجح لإعداد الحزم الاشتقاقية لجأ المختصون إلى تطوير المحللات الصرفية المتاحة، وتطوير أدوات حاسوبية أخرى لتحليل المدونة، واستخلاص الحزم الاشتقاقية، ومن ثمّ إحالتها إلى خبراء لغويين لاستكمال تهيئتها يدوياً وفق الضوابط المحددة، واستدراك الثغرات التي تركها المحلل الصرفي. وضمن الدقة يُحال العمل إلى خبراء لغويين لمراجعته)) (مؤسسة معجم الدوحة التاريخي، حول المعجم، 2021).

فمن ذلك البرمجيات المستخدمة في المعجم التفاعلي للغة العربية التي أعدت من طرف مدينة الملك عبد العزيز للعلوم والتقنية (KACST) بالاشتراك مع المنظمة العربية للتربية والثقافة والعلوم (ALECSO) والمعهد العالي للعلوم التطبيقية والتكنولوجيا (HIAST) (HIAST -ALECSO-KACST، 2021).

وكانت ولازالت أكبر المعضلات التي تواجه المعالجة الآلية للكلام هي تعرف البرمجية الدقيق على المدخل من خلال التعرف الآلي على شكله وسماته وموقعيته في سياق النص تمهيدا لتحديد دلالاته ومعناه الدقيق في ضوء هذه المعطيات. ذلك أن ((من أصعب نقاط البحث في المعالجة الآلية للنصوص فهم الدلالة، إذ قد يتعدد المعنى فيزداد اللبس بقدره)) (يوسف السبع، 2006، صفحة 71) وكان هذا من أكبر التحديات التي واجهت العاملين في مجال المعالجة الآلية للغة. وعليه، مرت المعالجة الرقمية لنصوص المدونة بعدد من المراحل والمستويات لتمكين وتسهيل التعرف الآلي على ألفاظ المدونة والتعرف عليها وهيكلتها ضمن نصوصها وسياقاتها المختلفة، تمهيدا للمعالجة المعجمية للمداخل ونصوصها¹. وفيما يأتي أهم هذه المراحل وفق ما أوردته مؤسسة معجم الدوحة التاريخي على موقع المعجم:

1.2.5.1.3. الترميز:

وهي عملية إجرائية يتم القيام بها من أجل توصيف نصوص المدونة لإتاحة التعرف الآلي عليها وتحليلها ((ولأنّ العربيّة تتمنّع بنظامٍ كتابيّ خاصٍّ بمحارفها، فقد تمّ ترميزُ وثائقِ المُدونة بصيغةٍ مُوحّدة، هي صيغة الترميز (CP-1256) التي تدعّمُ الألفبائيّة العربيّة، وتوجّهُ الآلةَ إلى فهم بنية اللّغة وتعيين أساليب مُعالجَتِها)) (مؤسسة معجم الدوحة التاريخي، حول المعجم، 2021).

2.2.5.1.3. الفهرسة الآلية:

¹ تمر المعالجة المعجمية للمداخل ونصوصها بعدد من المراحل أهمها:
تحديد السياقات المختلفة للمداخل المعجمية.
انقضاء الشواهد بناء على التطور التاريخي لمعنى الوحدة المعجمية بما يثبت ويوضح استعمالها ومعناها خلال المراحل الزمنية المختلفة.
ضبط المداخل ونصوص الشواهد ضبطا تاما والتأريخ لها.
توثيق نصوص الشواهد ونسبتها لأصحابها.
تحديد وسوم الوحدات اللغوية وصياغة تعريفاتها. (مؤسسة معجم الدوحة التاريخي، الدليل المعياري للمعالجة والتحرير المعجمي، 2021)

وهي عملية تعنى بحصر مفردات نصوص المدونة وتحديد سياقاتها المختلفة من أجل استخلاص بيانات إحصائية حولها واستكشاف الأخطاء تمهيدا لاستخلاص المباني من هذه المفردات والمعاني من السياقات من خلال المعالجة.

3.2.5.1.3. التحليل الصرفي:

وهو مرحلة من مراحل المعالجة الآلية للمادة اللغوية لنصوص المدونة تسمح لنا برد الوحدات الفروع إلى أصولها (الجزر المعجمي) واستخراج الوحدات الفروع للأصل الواحد، وهو ما يساهم في التوسيم الصرفي للمفردات اللغوية تمهيدا للمعالجة المعجمية على مستوى المباني والمعاني من خلال تهيئة الحزم الاشتقاقية المرتبطة بالأصول.

4.2.5.1.3. تهيئة الحزم الاشتقاقية:

يتم في هذه المرحلة المعالجة المعجمية من طرف الخبراء للمشتقات المرتبطة بأصل واحد وسياقاتها المختلفة المعبرة عن المراحل الزمنية للغة العربية ضمن ما يسمى بالحزم الاشتقاقية لاستخلاص المباني والمعاني وتبين مظاهر تطورها.

2.3. المعجم التفاعلي:

1.2.3. تعريفه:

1.1.2.3. تعريف المعجم الإلكتروني التفاعلي للغة العربية:

المعجم العربي الحاسوبي التفاعلي: وهو معجم عربي أحادي اللغة، يشتمل على إمكانية المعالجة اللغوية الآلية في عدة مستويات: مستوى المفردات والمستوى الصرفي والنحوي والدلالي والصوتي الإحصائي.

ويناسب هذا المعجم اللغة العربية لتمييزها بالنظام الاشتقاقي الصرفي والعلاقة بينه وبين النظم النحوية والدلالية (المنظمة العربية للتربية والثقافة والعلوم، مواصفات وخصائص المشروع، 2021).

2.1.2.3. تعريف معجم اللغة العربية التفاعلي للألكسو ALECSO:

هو معجم حاسوبي إلكتروني تفاعلي تم بناءه من طرف مدينة الملك عبد العزيز للعلوم والتقنية (KACST) بالاشتراك مع المنظمة العربية للتربية والثقافة

والعلوم (ALECSO) والمعهد العالي للعلوم التطبيقية والتكنولوجيا (HIAST)، يتميز هذا المعجم بخصائص برمجية فريدة، خاصة على مستوى برمجيات المعالجة الآلية للغة وإمكانات التحكم والتعديل على المحتوى، جعل المعجم يقدم محتوى معجمي حديث متجاوب مع مختلف فئات واحتياجات المستخدمين.

2.2.3. مميزات المعجم اللغوية والبرمجية

يتوافر المعجم على جملة من الخصائص والسمات التي تميزه عما سواه من المشاريع المعجمية في هذا الباب، والتي تتنوع بين مميزات لغوية وأخرى برمجية نورد أهمها في الآتي (المنظمة العربية للتربية والثقافة والعلوم، 2021):

1.2.2.3. الميزات اللغوية

- قدرته على استيعاب أكثر من 7000 جذر و 24000 فعل بين ثلاثي ورباعي، إضافة لاستيعابه لقرابة 14400 مصدر سماعي.
- تضمنه لنظام للتصريف والاشتقاق يعطيه المقدرة على توليد أكثر من 80000 مصدر قياسي واسم مشتق.
- احتواءه على مدقق إملائي إضافة لمحلل صرفي ونظام للتصريف والاشتقاق.
- استيعابه لقواعد رسم الهمزة في الأسماء والأفعال إضافة إلى القدرة على ضبط الكلمات بالشكل ضبطا تاما.

2.2.2.3. الميزات البرمجية

- إمكانية التعديل على المحتوى المعجمي لموقع المعجم إضافة وحذفا وتغييرا.
- استقلالية الموقع الإلكتروني للمعجم الحاسوبي التفاعلي عن البرنامج المثبت على الأجهزة.
- إتاحة القدرة على تطوير إمكانات المعالجة الآلية للمحتوى اللغوي لكونه مفتوح المصدر.

- توفير تجربة تصفح وإمكانات تحكم متميزة من خلال إتاحة خيارات متنوعة تسهل عملية الولوج للمحتوى المعجمي والتحكم في عرضه.
- تجاوب المعجم مع مختلف الأجهزة وعمله على مختلف أنظمة التشغيل.

الشكل 4: الواجهة الرئيسية للنسخة التجريبية من معجم اللغة العربية التفاعلي



3.2.3. برمجيات المعجم التفاعلي:

يتأسس عمل المعجم الحاسوبي التفاعلي على عدد من برمجيات المعالجة الآلية للغة، يمكن توزيعها من حيث طبيعة عملها والوظيفة التي تؤديها على قسمين، برمجيات تعمل مع محرك البحث وتعزز أداءه كمثل ما يؤديه كل من المدقق الإملائي، المشكل الآلي والمحلل الصرفي، وأخرى تعمل على مستوى المدخل المعجمي، بحيث تقدم معلومات ذات طابع لغوي عنه مثلما هو الأمر مع نظام التصريف والاشتقاق الذي تم بنائه تحت مظلة المنظمة العربية للتربية والثقافة والعلوم وفق مسمى "برنامج الخليل الصرفي"¹، وتدعم عملية المعالجة المعجمية لنصوص المعجم، ونعرف فيما يأتي بأهم هذه البرمجيات:

¹ نظام لاشتقاق وتوليد الأفعال والأسماء والمصادر وتصريفها، انطلاقاً من جذورها الثلاثية أو الرباعية، بالاعتماد على قواعد معطيات خاصة تتضمن قوانين النحو والصرف وعلى المعجم الحاسوبي لهذا النظام. تم إنجازه من قبل المنظمة العربية للتربية والثقافة والعلوم بالتعاون مع جامعة

1.3.2.3. المدقق الإملائي:

وتعد أولى البرمجيات التي تتفاعل مع مدخلات المستخدم وتقوم بتوجيهها وتصويبها، ويربط عملها بمحرك البحث والمحلل الصرفي، وتعد من البرمجيات التي لا يمكن الاستغناء عنها في المعاجم الإلكترونية والمدونات اللغوية الرقمية ومحركات البحث العالمية، ويعتمد عملها بالرجوع إلى قاعدة بيانات المدونة التي تتضمن كما ضخما من النصوص اللغوية المفهرسة والمجدولة لتسهيل القراءة الحاسوبية لها، والبحث عن أقرب خيار مطابق للمدخل أو اقتراح أقرب وحدة لغوية تتشابه مع المدخل، وبذلك فإن لهذه البرمجية دورا كبيرا في إرشاد المستخدم إلى الكتابة الصحيحة للمدخل وتسهيل توجيهه إلى ما يبحث عنه.

2.3.2.3. المشكل الآلي:

من أهم الخصائص المميزة للكتابة العربية اعتمادها على الشكل في تحديد المعنى الدقيق، إذ قد يغير تبديل حركة واحدة معنى الكلمة بشكل جذري، ومن هنا كان للمشكل الآلي أهمية كبرى ودور أساسي في المعجم التفاعلي، ويعمل هذا النظام من خلال التعرف الآلي على سياق الكلمة ضمن الجملة واقتراح خيارات الشكل المحتملة للكلمة ضمن السياق.

3.3.2.3. المحلل الصرفي:

وهو نظام تحليلي من أنظمة المعالجة الآلية للمدخلات اللغوية الغرض منه تحليل الكلمة لتحديد المعارف الصرفية المتعلقة بها كجذرها وسابقتها ولاحقتها ونوعها ووزنها، ويُعد هذا النظام من الأدوات الرئيسية المفتوحة المصدر في معالجة الآلية اللغة العربية، لذا فإنه يدخل في بنية الأنظمة الأخرى لمعالجة اللغة العربية، نذكر

محمد الأول بالمملكة المغربية ومدينة الملك عبد العزيز للعلوم والتقنية بالمملكة العربية السعودية عام 2010 (المنظمة العربية للتربية والثقافة والعلوم، برنامج الخليل الصرفي ، 2021).

منها المعجم العربي الحاسوبي التفاعلي (المنظمة العربية للتربية والثقافة والعلوم، برنامج الخليل الصرفي ، 2021).

وهو ما يجعل هذا النظام مرتبطا مباشرة بمحرك البحث من حيث مبدأ عمله، من خلال قيامه بدور تحديد المدخلات الأصول وتمييزها عن الوحدات الفروع وتحليل المدخلات التي لم يتم العثور على نتائج مطابقة لها على قاعدة البيانات، بتحديد الزوائد وفصلها عن الحروف الأصول، لتحديد أصل الكلمة، وربطه بما يقابله في صفوف قاعدة البيانات.

4.3.2.3. نظام الاشتقاق والتصريف:

وهو نظام لاشتقاق وتوليد الأفعال والأسماء المشتقة والمصادر وتصريفها انطلاقا من جذورها بالاعتماد على قواعد النحو الصرف.

وتعد قاعدة معطيات هذا النظام هي الأساس الذي يقوم عليه، بحيث تحوي ملفات تتضمن مختلف قواعد الاشتقاق والتصريف إضافة لقائمة للأسماء والأفعال وأوزانها يتم فيها تمييز أسماء العلم والمصطلحات (المنظمة العربية للتربية والثقافة والعلوم، نظام الاشتقاق والتصريف في اللغة العربية، 2021).

5.3.2.3. نظام البحث:

وهو بوابة الولوج للمحتوى وأداة استخراج ما تحتويه قواعد البيانات من بيانات لغوية، ويعد من أهم ما يميز المعجم الإلكتروني عن الورقي، والأداة التي تسهل على المستخدم الوصول للمحتوى اللغوي للمعجم، من خلال ما توفره من خيارات في الوصول للمحتوى وعرضه.

ولعله يصح الاعتقاد بأن محرك البحث . بالنسبة المعجم الحاسوبي . من أهم المكونات إن لم يكن أهمها على الإطلاق، ذلك أنه لا يمكن تسمية المعجم بالحاسوبي إذا لم يحتوي على محرك بحث.

4.2.3. آلية عمل معجم اللغة العربية التفاعلي للألكسو ALECSO:

تتحدد آلية عمل معجم اللغة العربية التفاعلي من خلال تضافر عمل عدد من الطبقات البرمجية، وتم الاعتماد في بناء المعجم على البنين ذي الطبقات الأربع، وتم بناء على ذلك توزيع الصفوف فيما بين طبقة العرض والتمثيل الخارجي Presentation Layer: وهي الطبقة المتعلقة بطلب البيانات وعرضها والتي يتعامل معها المستخدم مباشرة من خلال المستخدم، وطبقة منطق العمل Business Logic Layer: وهي الطبقة التي تربط بين طبقة العرض وطبقة الوصول للمعطيات وتتضمن أغلب تعليمات عمل النظام ووظائفه، وطبقة النفاذ للمعطيات Data Access Layer: والمتعلقة بتأمين الاتصال بقاعدة المعطيات، وطبقة المعطيات الدائمة Persistence Layer: وهي بمثابة واجهة ربط بين التطبيق وقاعدة المعطيات (غنيم وآخرون، 2013، الصفحات 203-204).

5.2.3. المعالجة الآلية للمادة المعجمية

ويبدأ ذلك بضبط المدونة اللغوية، وتحديد المصادر التي سيعتمد عليها المعجم في جمع مادته واستقاء نصوصه وشواهد، بحيث تضمّ كما وافية من النصوص التي استقت مادتها من مصادر اللغة العربية المختلفة (مجموعة من المؤلفين، 2014، صفحة 153).

لتخضع المدونة في مرحلة لاحقة للمراجعة والتنقيح، ويشترط في هذه المدونة أن تواكب التطور الحاصل للغة وما استجد فيها من ألفاظ ومصطلحات.

ويمكن الاستعانة في بناء المدونة ببعض البرمجيات الحرة المستخدمة في هذا المجال، كنظام NOOJ الذي تم الإشارة إليه في مقترح الأدوات والخبرات المستخدمة لبناء المعجم التفاعلي في مشروع إعداد المعجم (HIAST-ALECSO-KACST، معجم اللغة العربية التفاعلي: التصميم المفاهيمي للمشروع، 2021).

وتعد المدونة المصدر الأساس لمادة المعجم، بالنظر إلى ما توفره من ألفاظ ونصوص شواهد، إضافة لدورها في عمل برمجيات المعالجة الآلية للغة.

بعد الانتهاء من إعداد المدونة يتم ضبط قائمة المداخل الأول والوحدات اللغوية التي تتبع لها، إضافة لتحديد نصوص الشواهد المرتبطة بهذه المداخل.

وتتحدد المرحلة اللاحقة في صياغة تعاريف وشروح المداخل من طرف المحررين المعجميين، وتحديد المداخل التي تحتاج لإضافة محتوى مرئي أو مسموع.

تمر عملية طلب المعلومة من خلال محرك البحث بعدد من عمليات المعالجة ضمن عدد من المستويات تتولى تنفيذها برمجيات المعالجة اللغوية بالمعجم.

فأثناء قيام المستخدم بطلب بحث عن طريق كتابة مدخل ما مباشرة يقوم المعجم بالبحث عما يقابل المدخل في قاعدة البيانات، فإن لم تكن هناك نتيجة، يبدأ عمل المدقق الإملائي الذي يقوم بمحاولة تفسير الأخطاء الإملائية التي من الممكن وقوعها ليحيل المستخدم إلى الخيارات الأقرب لتصحيح نص البحث، أو يتم تقديم بدائل عن طريق اقتراح تعديل، بناء على المقابلة بين المدخل المقترح وما يقابله من نتائج مشابهة في ملف قاعدة بيانات يحتوي على جميع المداخل المعجمية الواردة في المعجم ليتم إحالة المستخدم لبدائل أخرى لها علاقة بالمدخل الذي أراد البحث عنه (تاويريت، 2020، صفحة 304).

في المرحلة الموالية يتم استخدام نظام المحلل الصرفي في استخراج أصل الكلمة وذلك بتجريد الكلمة من الزوائد وإرجاعها إلى حروفها الأصول، لتسهيل عملية التعرف على الكلمة ومن ثمة تحديد معناها، ليتم عرض النتيجة النهائية للمستخدم.

إن مدى نجاح المعجم الإلكتروني في أداء مهامه يقاس بقدرته على إحالة المستخدم إلى اللفظ الذي يطابق المعنى الذي في ذهنه أو المعنى الدقيق الذي يطابق اللفظ الذي في ذهنه بإيجاز وسهولة، مع الاستعانة عند الضرورة ببعض محددات دعم المعنى والاستفادة من برمجيات المعالجة الآلية للغة في تحقيق ذلك.

4.3. المدونة اللغوية العربية العالمية

1.4.3. تعريفها

المدونة اللغوية العربية العالمية واحدة من المحاولات اللغوية العربية لبناء مدونة لغوية إلكترونية للغة العربية المعاصرة تشرف عليها مكتبة الإسكندرية.

تحتوي المدونة حوالي 100 مليون كلمة محللة صرفيا ونحويا ودلاليا، تشمل قطاعا إقليميا واسعا من الدول الناطقة بالعربية بحيث تعكس واقع استعمال اللغة العربية المعاصرة في أنحاء هذه الدول (مكتبة الإسكندرية، 2021).

3.4.3. مصادر نصوص المدونة وتوزعها الجغرافي:

تم الاعتماد في جمع نصوص المدونة على أربعة مصادر أساسية: الصحافة والمقالات الإلكترونية والكتب والدراسات الأكاديمية، وقد احتوى كل مصدر من هذه المصادر على مجموعة من الفئات التي تعبر عن التنوع اللفظي والمعرفي لنصوص المدونة.

أما بالنسبة للتوزيع الجغرافي لمصادر الجمع فإننا نلاحظ وجود نوع من التوازن في توزيع عدد كلمات نصوص المدونة بين دول الوطن العربي بما يكفل التنوع والثراء في نصوص المدونة.

وكما يظهر من خلال الشكل البياني فإن المدونة تغطي جميع المنشورات داخل الوطن العربي وكذلك بعض المنشورات العربية المنشورة خارج الوطن العربي.

4.4.3. التحليل الصرفي لنصوص المدونة اللغوية العربية العالمية

يجري حاليا التحليل الصرفي لكلمات المدونة باستخدام المحلل الصرفي تيم باك والتر Tim Buckwalter، لاستخراج المعلومات الصرفية والإحصائية المتعلقة بالكلمات: كجذر الكلمة، السوابق واللواحق، الوزن الصرفي، الجنس، العدد، وإتاحة محتوى المدونة للباحثين للاستفادة منها في عمل دراسات وأبحاث، وتصميم برمجيات لتطوير ضمن المحاولات الساعية لتطوير مجال المعالجة الآلية لنصوص اللغة العربية.

ومن الممكن أن يكون هذا المشروع الطموح مستقبلا؛ أساسا لمشاريع معجمية تخدم اللغة العربية وتساهم في تطويرها وتقريبها من القارئ والمتعلم (مكتبة الإسكندرية، 2021).

4. خاتمة

بعد الاطلاع على بعض من أهم التجارب العربية في رقمنة اللغة العربية ومعالجتها آليا في إطار مشاريع لغوية ومعجمية متنوعة، توصلنا إلى عدد من النتائج حول هذا الموضوع يمكن إجمالها في الآتي:

. تعد المدونات اللغوية أرضا خصبة لإقامة مشاريع معجمية ولغوية متنوعة.

. يرتبط عمل كثير من برمجيات المعالجة الآلية للغة بالمدونات اللغوية .

. يعد مشروع بناء مدونة لغوية إلكترونية بمثابة اللبنة الأولى والحجر الأساس

لإقامة مشروع المعجم الحاسوبي للغة العربية.

. يمر إعداد المدونات اللغوية الإلكترونية بعدد من المراحل أهمها: تحديد

نصوص المدونة ومصادرها . نقل نصوص المدونة من الحالة الورقية إلى الإلكترونية .

تدقيق نصوص المدونة ومراجعتها لغويا . توصيف نصوص المدونة وإعدادها للقراءة

الحاسوبية تمهيدا للمعالجة الآلية.

. تستقي المعاجم الإلكترونية مادتها وشواهداها من المدونات اللغوية، وتعتمد في

آلية عملها وعمل برمجياتها على هذه المدونات.

. يعد مشروع معجم الدوحة التاريخي تجربة رائدة في مجال الاستفادة من

المدونات اللغوية في صناعة معجم حاسوبي تفاعلي تاريخي.

. يظهر بأن مشروع المعجم التفاعلي للغة العربية للألكسو قد قطع أشواطاً في

مجال المعالجة الآلية للغة والمعالجة المعجمية الآلية والاستفادة من ذلك بناء المعجم

وتطويره.

. استفادت مؤسسة معجم الدوحة التاريخي مما سبق من جهود في مجال رقمنة

المحتوى العربي وبرمجيات سابقة في مجال التحليل الصرفي والتدقيق الإملائي، ما

كان له دور كبير في توفير الوقت والجهد والمال.

. تستلزم المعالجة الآلية في المعجم الإلكتروني توفر عدد من البرمجيات أهمها:

مدقق إملائي . محلل صرفي . نظام للاشتقاق . محرك بحث.

قائمة المراجع

- . HIAST –ALECSO–KACST. (2021, 02 24). معجم اللغة العربية التفاعلي: التصميم المفاهيمي للمشروع. الموقع الرسمي للمنظمة العربية للتربية والثقافة والعلوم : [.https://ossl.alecso.org/affich_oso_details.php?id=56](https://ossl.alecso.org/affich_oso_details.php?id=56)
- . HIAST –ALECSO–KACST. (2021, 02 03). معجم اللغة العربية التفاعلي: التقرير الفني. المنظمة العربية للتربية والثقافة والعلوم : [.https://ossl.alecso.org/affich_oso_details.php?id=56](https://ossl.alecso.org/affich_oso_details.php?id=56)
- . الجمعاوي, ص. (2013). المعالجة الآلية للمصطلح الطبي العربي. مجلة التعريب. 15, (44). المنظمة العربية للتربية والثقافة والعلوم. (2021, 02 23). المعجم الحاسوبي التفاعلي للغة العربية. الموقع الرسمي للمنظمة العربية للتربية والثقافة والعلوم : [.https://ossl.alecso.org/affich_oso_details.php?id=56](https://ossl.alecso.org/affich_oso_details.php?id=56)
- . المنظمة العربية للتربية والثقافة والعلوم. (2021, 02 23). برنامج الخليل الصرفي. الموقع الرسمي للمنظمة العربية للتربية والثقافة والعلوم : [.https://ossl.alecso.org/affich_oso_details.php?id=57](https://ossl.alecso.org/affich_oso_details.php?id=57)
- . المنظمة العربية للتربية والثقافة والعلوم. (2021, 02 21). مواصفات وخصائص المشروع. الموقع الرسمي للمنظمة العربية للتربية والثقافة والعلوم: www.alecso.org
- . المنظمة العربية للتربية والثقافة والعلوم. (2021, 02 24). نظام الاشتقاق والتصريف في اللغة العربية. الموقع الرسمي للمنظمة العربية للتربية والثقافة والعلوم : [.https://ossl.alecso.org/affich_oso_details.php?id=55](https://ossl.alecso.org/affich_oso_details.php?id=55)
- . بسو, ص. (2016). المعالجة الآلية للغة العربية في الإدارة الإلكترونية. ندوة حول اللغة العربية وتحديات الإدارة الإلكترونية. الجزائر: المجلس الأعلى للغة العربية. (ص.5).
- . بن حامدو, ع. (2011). المعجم العربي الإلكتروني: أهميته وطرق بنائه. صفاقص: جامعة صفاقص.
- . تاويريت, ح. (2020). المعجم التفاعلي للغة العربية: أدواته المنهجية وخياراته الإجرائية. مجلة علوم اللغة العربية وآدابها, 12(1), 304.
- . صالح, م. (2021, 02 02). لسانيات المدونات اللغوية : مقدمة للقارئ العربي. مدونة الدكتور محمود إسماعيل صالح: <http://dr-mahmoud-ismail-saleh.blogspot.com/>
- . غنيم, ن. و آخرون. (2013). التشاركية في إغناء معجم اللغة العربية التفاعلي. مجلة جامعة دمشق للعلوم الهندسية, 29(1), 204-203.

- . مجموعة من المؤلفين .(2014) نحو معجم تاريخي للغة العربية . بيروت: المركز العربي للأبحاث ودراسة السياسات.
- . مدينة الملك عبد العزيز للعلوم والتقنية .(2021, 02 24) .المدونة اللغوية العربية .الموقع الرسمي للمدونة اللغوية العربية: <https://corpus.kacst.edu.sa> .
- . مكتبة الإسكندرية .(2021, 02 25) .المدونة اللغوية العالمية .الموقع الرسمي للمدونة اللغوية العالمية <https://www.bibalex.org/ica> .:
- . مؤسسة معجم الدوحة التاريخي .(2021, 02 24) .الدليل المعياري للمعالجة والتحرير المعجمي . الموقع الرسمي لمعجم الدوحة التاريخي: <https://dohadictionary.org/standard-guide> .
- . مؤسسة معجم الدوحة التاريخي .(2021, 02 24) .حول المعجم .الموقع الرسمي لمعجم الدوحة التاريخي <https://dohadictionary.org/about-dictionary> .:
- . مؤسسة معجم الدوحة التاريخي .(2021, 02 24) .حول المعجم .الموقع الرسمي لمعجم الدوحة التاريخي <https://dohadictionary.org/about-dictionary> .:
- . مؤسسة معجم الدوحة التاريخي .(2021, 02 24) .عن المدونة اللغوية .الموقع الرسمي لمعجم الدوحة التاريخي <https://dohadictionary.org/corpus> .:
- . مؤسسة معجم الدوحة التاريخي .(2021, 02 24) .عن المعجم .الموقع الرسمي لمعجم الدوحة التاريخي <https://news.dohadictionary.org/AR/AboutUs/Pages/default.aspx> .:
- . مؤسسة معجم الدوحة التاريخي .(2021, 02 24) .عن مصادر المعجم .الموقع الرسمي لمعجم الدوحة: التاريخي <https://dohadictionary.org/bibliography> .
- . يوسف السبع، م .(2006) .من وسائل فك اللبس الدلالي للمعالجة الآلية للعربية(نموذج الفعل) . مجلة المعجمية. 71، (21-22).