

دور المدونات الحاسوبية في بناء إقتصاد المعرفة -التجربة العربية أنموذجا-

د.أنوار بن شوك *

—مركز البحث العلمي والتقني لتطوير اللغة العربية —وحدة تلمسان

د. نادية سعدي

—جامعة ابن خلدون تيارت

تاريخ الارسال 11-03-2024 تاريخ القبول 06-05-2024 تاريخ النشر 31-12-2024

الملخص :

المدونات أو الذخائر أو المتون اللغوية الحاسوبية هي نصوص إلكترونية ضخمة؛ جمعت لتكون ممثلة للواقع اللغوي من خلال تمكين اللغة العربية في مجتمع إقتصاد المعرفة، على الصعيد العلمي و الاجتماعي، وعلى الرغم من أن التجربة العربية في مجال المعالجة الآلية لاتزال في بداياتها، إلا أن التقدم الهائل في مجال اللغويات الحاسوبية يحاصر استرخاءنا اللغوي والتقني في زاوية ضيقة، فقد برزت الدوافع، وتوفرت الأساليب اللازمة لوضع علاقة العربية بالحاسوب تحت الفحص المجهرى الدقيق، لضمان بقاء هاته اللغة، وعليه جاءت هذه الدراسة لتبين دور المدونات الحاسوبية في بناء إقتصاد المعرفة ملقن الضوء على أهم التجارب العربية التي كان لها الدور الفاعل في ذلك، لتكون الإجابة تترجم التساؤل التالي: ما واقع المدونة العربية في مجال برمجة الأنساق اللغوية؟ وما الدور الذي اضطلعت به في مجال إقتصاد المعرفة؟

الكلمات المفتاحية: لسانيات المدونات ، المدونات اللغوية، الأنسقة اللغوية، الواقع اللغوي، إقتصاد المعرفة.

Abstract:

Codes, munitions or computer language tools are large electronic texts; Although the Arab experience of automated processing is still in its infancy, the tremendous progress in computer linguistics is besieging our linguistic and technical relaxation at a narrow angle. The motives have emerged, and the necessary methods are in place to place Arabic's relationship with the computer under careful microscopy to ensure that this language remains. What's the reality of the Arabic blog in language programming? What role has she played in the knowledge economy?

* المؤلف المرسل: أنوار بن شوك benchouk.anouar@gmail.com

Keywords: blog languages, language blogs, language textures, linguistic reality, knowledge economy.

1. مقدمة:

جدير بالذكر أنّ الإقتصاد العالمي منذ ثمانينات القرن العشرين تحوّل من إقتصاد صناعي إلى ما يعرف بإقتصاد المعرفة؛ وذلك بصفته نتيجة طبيعيّة للإستثمارات الكبيرة في مجالات البحث العلمي والتطبيق. الأمر الذي جعل المنتج الإستراتيجي الأبرز لأيّ أمةٍ هو المعرفة التي لا يمكن تناقلها إلّا من خلال لغة تحمل رسائلها عبر الأجيال والثقافات المختلفة.

ومنه فقد وُفّرت التكنولوجيا الحديثة وعلى رأسها الإنترنت، في تسعينات القرن الماضي الأرض الخصبة لنمو إنتاج اللّغة بمعدلات مطّردة غير معهودة، حتى أصبحت السوق هي العالم بأسره، ممّا أدى إلى إنتاج عشرات الآلاف من الوثائق التي تحمل لغات العالم في كلّ دقيقة، وقد صاحب ذلك كلّ إضافة قدرٍ من التعقيد على الصورة التي توجد عليها نصوص هذه اللّغات، وجعلها متنوّعة بشكل كبير، وهكذا أضيفت أبعاد جديدة وتبعات كثيرة، وأستحدثت أنماط من مصادر معلومات لم نألّفها من قبل فأصبح الباحث تائها في خضمّ هذا الزحام وهذا التنوّع، وفي ظلّ هذا الظرف تنبّه العالم إلى مخاطر اتّساع الفجوة بين فيض النصوص و المعلومات، وبين إمكانيّة تنظيمه و التحكم فيه، وهي فجوة تعاني منها الكثير من لغات العالم، ولاسيّما اللّغة العربيّة، حتّى باتت قضية تطويع التكنولوجيا الحديثة لخدمة اللّغة العربيّة واحدة من أهمّ القضايا الأكثر إلحاحا، والتي يتحتّم تناولها ومعالجتها على التّحو الذي من شأنه أن يرأب بين ما يتّم إنتاجه من لغةٍ، وبين إحكام السيطرة عليه، وباعتبارنا أبناء اللّغة العربيّة يتعيّن علينا المحافظة عليها بكافة السبل الممكنة، حتى تظلّ قويّة في مواجهة تحديات و صراعات العولمة، ومن أبرز السبل للحفاظ عليها رصد و استكشاف كلّ ما من شأنه أن يحفظ و يعالج و يتيح لغتنا —بما تحويه من نصوص و معلومات على اختلاف مستوياتها و أنواعها— بالسرعة و السهولة المطلوبتين، ومن هنا دعت الحاجة إلى تطويع التكنولوجيا الحديثة للمعالجة و التطبيق مستفيدة ممّا توفره إمكانيات العصر من تقنيات و برمجيات.

وتعدّ المدونات اللّغويّة Corpora أحد السبل التي إستفادت كثيرا من التكنولوجيا الحديثة، وعلى الرغم من أنّ فكرة المدونات اللّغويّة قائمة منذ آلاف السنين، إلّا أنّ التكنولوجيا الحديثة أحيّتها من جديد، وأصبح هناك إهتمام متزايد في الآونة الأخيرة بمثل هذا المصدر من مصادر المعلومات، فبعد أن كانت عمليات إنشاء وإتاحة

المدونات اللغوية تشكّل عبئا وجهدا كبيرين- لما تتطلبه من فرق عمل، إضافة إلى الصعوبة البالغة في جمع البيانات الخام من آلاف النصوص- أصبحت هذه العمليات سهلة المنال نسبيا بفضل التوجه السائد نحو رقمنة نصوص مصادر المعلومات الأولية و نشرها كي تصبح جزءا من قواعد البيانات ومحتوى الشبكة العنكبوتية.

وعليه نسعى من خلال هذا العمل أن نقف عند أهمّ النماذج التطبيقية للمدونات اللغوية العربية الحاسوبية، وتبيان دورها الفاعل في بناء إقتصاد المعرفة، محاولين الإجابة عن الإشكالية الآتية:

- ما واقع المدونة العربية في مجال برمجة الأنساق اللغوية؟ وما الدور الذي اضطلعت به في مجال إقتصاد المعرفة؟

1.1. أهداف الدراسة :

إنّ الدراسة -التي نحن بصدد تقديمها- نستقي أهدافها من معين الموضوع نفسه و أهميته، ؛حيث تكمن في النقاط الآتية:

- الوقوف على أهمّ التجارب العربية في مجال حوسبة المدونات اللغوية.
- ما تقدّمه المدونات اللغوية الحاسوبية-العربية- في مجال إقتصاد المعرفة.
- رصد أهمّ العوائق والتحديات في سبيل تطويع التكنولوجيات الحديثة نحو خدمة المدونات اللغوية العربية الحاسوبية.

2.1. مصطلحات الدراسة:

1.2.1. المدونة اللغوية:

أ- المعنى اللغوي المعجمي:

كلمة ((Corpus)) كلمة لاتينية تعني الجسد body، وجمعها Corpora أو Corpuses¹، ولكلمة Corpora مقابلات عدّة في العربية منها: المدونات اللغوية، والذخائر النصية، و الذخائر اللغوية، و المكتنزات النصية، و المتون اللغوية، وغيرها ممّا ذكر في مصادر غير أكاديمية، نحو: المفهرسات، المجموعات النصية، المخزونات النصية... إلخ².

ب- المعنى الاصطلاحي:

" مجموعة من النصوص اللغوية الآلية محدّدة الحجم مكتوبة كانت أو منسوخة صوتيا، تمّ جمعها بطريقة مدروسة وليس بشكل عشوائي، ممثلة للمجالات التي تمّ جمعها بهدف دراسة ظاهرة لغوية معيّنة أو بهدف الدراسة اللغوية بصفة عامة."³

المدوّنات اللغوية حسب الدكتور أيمن الدكروزي: " عبارة عن رصيد ضخّم من النصوص المكتوبة أو المنطوقة التي يتمّ اختزانها على الحاسب الآلي."⁴

من التعريفين السابقين يتبيّن لنا الشروط الواجب توافرها في المدوّنة:

- 1- أن تكون نصوصا مكتوبة أو نسخا صوتيا لكلام منطوق تمّ تسجيله مسبقا.
- 2- أن تكون موجودة على شكل آلي (machine-readable) حتى يسهل معالجتها فيما بعد.
- 3- أن تكون محدّدة الحجم (Finite size)*.
- 4- أن تكون مجموعة بطريقة مدروسة، حسب الغرض الذي جمعت لأجله، وذلك بأن تكون ممثلة للمجالات والمصادر التي جمعت منها، ومتوازنة بحيث تطابق الواقع من حيث نسب المجالات المكوّنة لها.
- 5- أن تكون متاحة للباحثين (standard reference) المعنيين بدراسة اللغة بجميع مستوياتها.

2.2.1. إقتصاد المعرفة:

يعرّف إقتصاد المعرفة بأنّه: " القدرة على إنتاج المعرفة و المعلومات و استخدامها و نشرها، و بناء معارف أكاديمية عن طريق وسائل و تقنيات جديدة تعتمد على التراكم المعرفي و المهارات التي يمتلكها العنصر البشري."⁵

ويعرّفه "محمد إبراهيم" علي بأنّه: " نمط إقتصادي متطوّر قائم على الاستخدام الواسع لتكنولوجيا المعلومات، والاتّصالات في مختلف أوجه النشاط الاقتصادي، والذي تمثّل فيه المعرفة عنصر الإنتاج الرئيس، والقوّة الدافعة لتكوين الثروة، ويهدف إلى التحسن في واقع الأداء الاقتصادي لتحقيق تنمية اقتصادية و بشرية مستدامة."⁶

نستنتج من التعريفين السابقين أنّ إقتصاد المعرفة يتشكل بالاعتماد على التكنولوجيات الحديثة في استثمار القوى البشرية لتحقيق التنمية المستدامة.

أما إجرائيًا: الإقتصاد الذي يعتمد على تحويل المعرفة المتعلقة باللغة في مختلف مجالاتها إلى ثروة مستدامة من خلال رصدها، وتخزينها وكنزها في بنك آلي.

3- بناء المكانز و تحليل المدونات:

كان للباحثين المصريين النصيب الأوفر من دراسات بناء المدونات اللغوية وتحليلها، ففي عام 2007، "صمم المعتر بالله طه مدونة لغوية إلكترونية بهدف بناء معجم يعكس الواقع اللغوي للغة العربية في الوقت الراهن، وبعدها بثلاث سنوات قدّم الباحث نفسه في أطروحته لنيل درجة الدكتوراه هيكلًا لمدونة المعجم التاريخي للعربية في محاولة لإحياء المعجم التاريخي، فطوّر مدونة لغوية لمعجم تاريخي للغة العربية، وفي عام 2016 سعى أحمد روبي إلى بناء مدونة لغوية للغة العربية الفصحى المعاصرة، وهيئة مادتها للغة العربية، كما قدّم منهاجًا لبناء بنك شجري نحوي للغة العربية الفصحى المعاصرة في ضوء المدونة اللغوية، وفي عام 2013 أعدت الباحثة نقاز هجيرة، جامعة أبو بكر بلقايد، تلمسان، الجزائر، مدونة لغوية حاسوبية تمثل أسماء المواقع الجغرافية في مدينة تلمسان.⁷

4- المدونات اللغوية: - نماذج عربية -

نحاول في هذا العنصر التطرق إلى مجموعة من المدونات اللغوية القائمة على الشبكة العنكبوتية، مع تسليط الضوء على العربية منها من حيث حجمها، و المجالات و الموضوعات التي تغطيها، و سماتها و خصائصها، و إمكانات البحث التي تقدّمها.

1.4. المدونة اللغوية العربية لمدينة الملك عبد العزيز للعلوم و التقنية:

1.1.4. التعريف بالمدونة:

هي مدونة لغوية عربية متاحة بالبحر من خلال هذا الرابط: corpus.kacst.edu.sa/index.jsp⁸، "هذه المدونة اللغوية لها أغراض بحثية مختلفة؛ تبدأ بالدراسات اللغوية بمستوياتها المتنوعة، وتمتد لتشمل تطوير تطبيقات معالجة اللغة الطبيعية."⁹

أنشئت هذه المدونة عام 2012، ويتجاوز حجمها المليار كلمة (182515633 كلمة)؛ وبذلك فهي تعدّ واحدة من أكبر المدونات العربية، كما أنّها تمثّل النصوص المكتوبة للفصحى، دون اللهجات العربية المختلفة،

وذلك عبر أبعاد رئيسية، هي: الفترة الزمنية، والمنطقة الجغرافية، والتنوع الأدبي، وتمتد الفترة الزمنية المغطاة من عصر ما قبل الإسلام، وحتى وقت إنشاء هذه المدونة اللغوية.

تمّ جمع نصوص المدونة من مصادر معلومات متنوعة، اعتماداً على ما هو متوافر على الشبكة العنكبوتية من محتوى عربي، وتشمل 10 مصادر معلومات، وهي: المخطوطات المحققة، والكتب، والصحف، والمجلات، والمناهج الدراسية، والرسائل العلمية، والمواقع الإلكترونية، والدوريات المحكمة، والمطبوعات الرسمية، ووكالات الأنباء، حيث تمّ تصنيف هذه النصوص على 80 نطاقاً، و 481 موضوعاً.¹⁰

الشكل رقم (01) الصفحة الرئيسية للمدونة اللغوية العربية لمدينة الملك عبد العزيز للعلوم والتقنية.

<https://corpus.kacst.edu.sa/index.jsp>

2.1.4. مميزات المدونة اللغوية العربية لمدينة الملك عبد العزيز للعلوم والتقنية :

تتميّز المدونة اللغوية العربية لمدينة الملك عبد العزيز للعلوم والتقنية بجملة من المميزات أهمّها¹¹:

- تعزيز نصوصها بالميتاداتا وفق حقول الوصف: العنوان، وتاريخ النشر، والفترة الزمنية، واسم المؤلف وجنسه والمنطقة، ومصدر اللغة، والنطاق، والموضوع، الأمر الذي يمكن الباحثين والمستفيدين من إمكانية الاسترجاع والتحليل وفق مقيدات بحث مختلفة، حيث توفر عددا من إمكانيات الاسترجاع، وهي كالاتي:

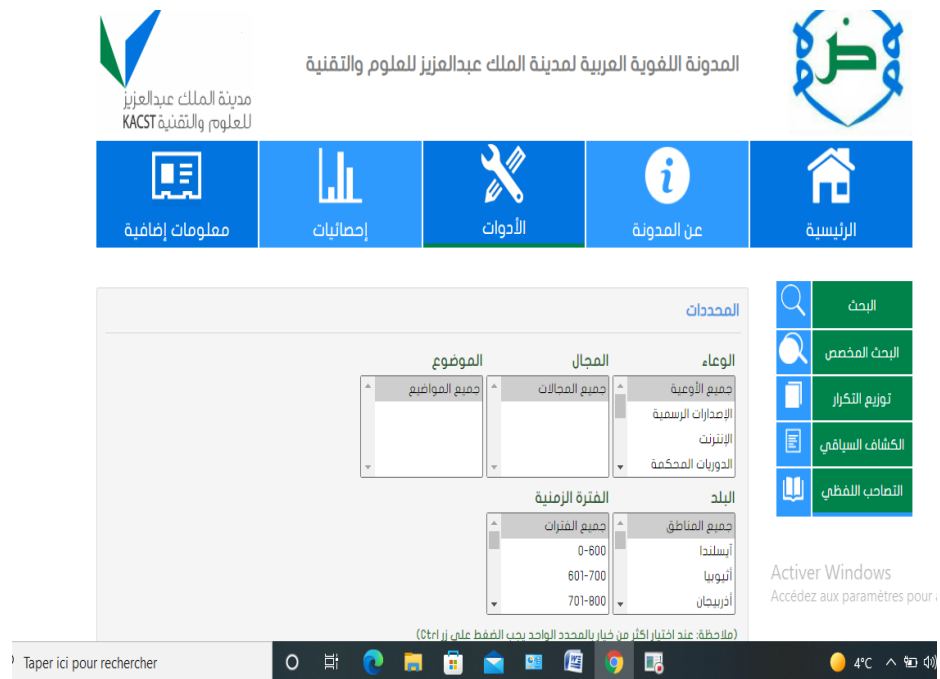
1.2.1.4 البحث:

توفر هذه الأداة نوعين من البحث¹²:

أ- **البحث في نصوص المدونة** بكلمة واحدة أو عدّة كلمات بطريقة مشابهة لمحركات البحث، حيث تظهر السياقات التي وردت فيها هذه الكلمات مع معلومات تحوي عنوان النص، والوعاء، والمجال، والموضوع، والفترة الزمنية، والبلد الذي طبع فيه النص، ويمكن للمستفيد تحديد البحث في جزء أو عدّة أجزاء محدّدة من المدونة من خلال المحدّات الموجودة أعلى الصفحة، الأمر الذي يمكّننا من تحديد أكثر من عنصر من عناصر المحدّد الواحد بالضغط على الزر CTRL من لوحة المفاتيح واختيار العناصر المطلوبة.

إمكانية البحث عن عناوين النصوص بواسطة كلمة أو مجموعة كلمات، وفيما يلي الشكل الذي يوضّح ذلك :

الشكل رقم (02): يوضح صفحة البحث في المدونة اللغوية العربية لمدينة الملك عبد العزيز للعلوم والتقنية.



<https://corpus.kacst.edu.sa/index.jsp>

ب- البحث المتخصص:

تمكّن هذه الأداة المستفيد من البحث عن الكلمات بواسطة: الجذع؛ بحيث يظهر الجذع بصوره المختلفة بعد إضافة السوابق و اللّواحق الممكنة مثل (أل التعريف، الواو، الباء، و لضمائر المتّصلة) و ذلك بحسب نوع الجذع (اسم، فعل)، مع معلومات عن تكرار هذه الصّور في المدوّنة، ويمكن من خلال النقر على الكلمة الانتقال إلى الكشف السياقي للكلمة.¹³ وفيما يلي شرح مبسط لأهم الرموز المعتمدة في المدوّنة:

- (/) و تعني أي عددٍ من الحروف.

- (-) وتعني حرفا واحدا فقط.

توضيح: يمكن أن يوضع أي رمز من هذين الرمزين (_)، (/) في أي مكان من سلسلة الحروف التي نود البحث عنها، فعند إدخال السلسلة (م /) فيعني هذا البحث عن أي كلمة فيها حرف الميم، وتعني السلسلة (م /) أي كلمة أولها حرف الميم، وتعني السلسلة (- م) البحث عن أي كلمة مكوّنة من ثلاثة أحرف وسطها حرف الميم. ويوضّح الشكل الآتي صفحة البحث في هذه المدونة:

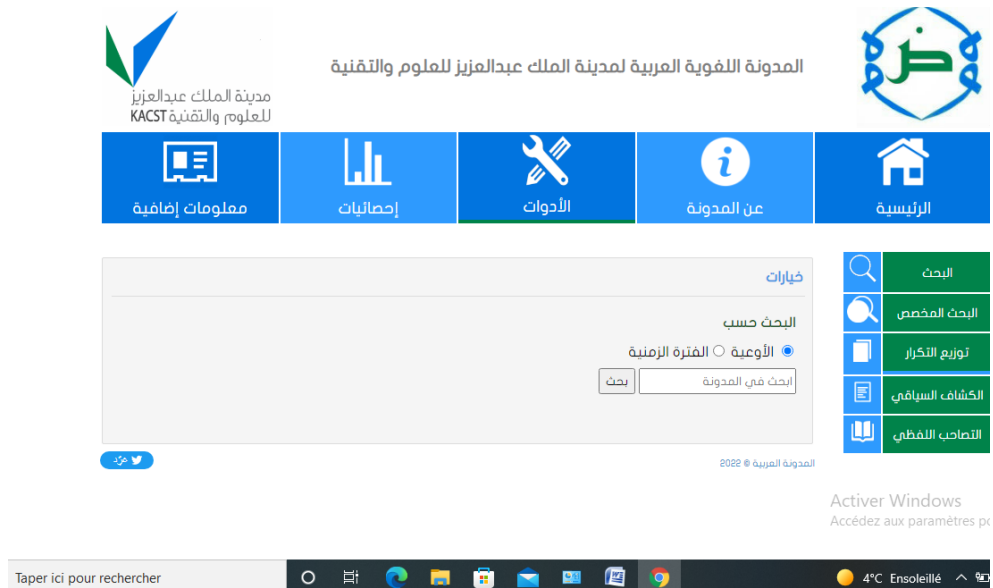
الشكل رقم (03): صفحة البحث المخصّص في المدوّنة اللّغويّة العربيّة لمدينة الملك عبد العزيز للعلوم و التقنيّة.

https://corpus.kacst.edu.sa/word_frequency.jsp

ج- توزيع التكرار:

يمكن من خلال هذه الأداة معرفة التوزيع الإحصائي للكلمة الواحدة، أو لكلمتين متتابعين، أو ثلاث كلمات حسب الفقرات الزمنية، وما يختاره الباحث، حيث تظهر الفترة الزمنية ويقابلها تكرار الكلمة، وتكرارها النسبي، وعدد النصوص التي وردت فيها. ويظهر الشكل الآتي صفحة توزيع التكرارات في المدونة:

الشكل رقم (04) يوضح صفحة توزيع التكرار في المونة اللغوية العربية لمدينة الملك عبد العزيز للعلوم والتقنية.

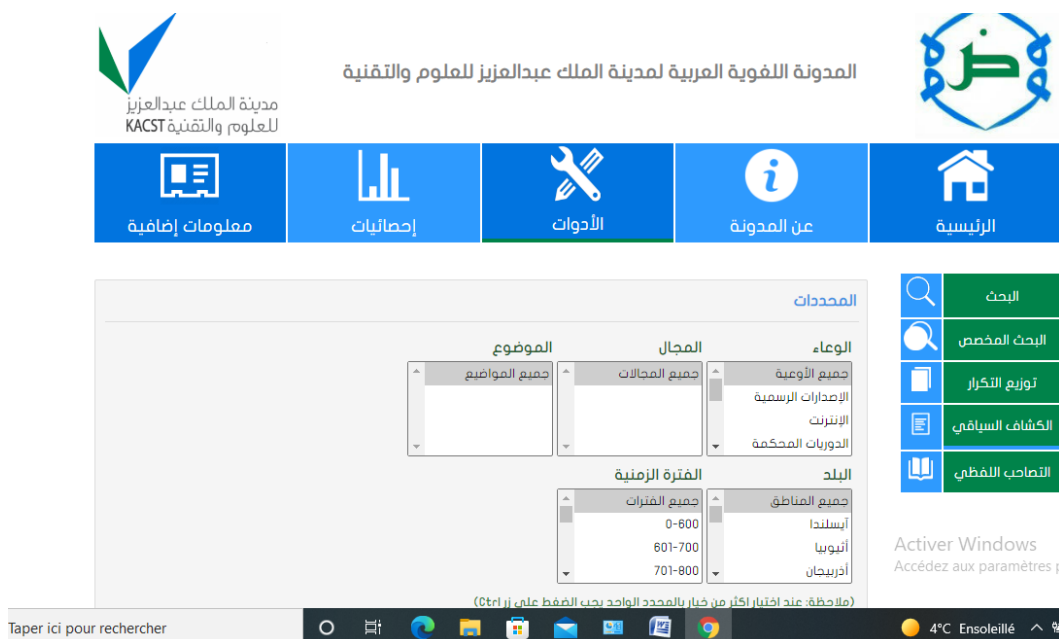


https://corpus.kacst.edu.sa/word_frequency.jsp

د- الكشاف السياقي:

تستخدم هذه الأداة لكشف و عرض السياقات التي تظهر فيها الكلمة، أو أي مجموعة من الكلمات المتتابعة ضمن مدى محدد من الكلمات السابقة و اللاحقة، ويتراوح عدد ها من صفر (0) و حتى (15) كلمة، كما وتكشف هذه الأداة عن دلالة الكلمة المركزية للسياق، وجدير بالذكر أنه يمكن للمستفيد حفظ هذه النتائج في حاسبه الآلي من خلال الضغط على زر الحفظ الموجود أعلى الصفحة أو أسفلها، والشكل الآتي يوضح صفحة الكشاف السياقي في المدونة.¹⁴

الشكل رقم (05): صفحة الكشف بالمدونة اللغوية العربية لمدينة الملك عبد العزيز للعلوم والتقنية.



<https://corpus.kacst.edu.sa/collocation.jsp>

هـ- التصاحب اللفظي:

تستخدم هذه الأداة لقياس مدى الارتباط بين الكلمة المبحوث عنها، و الكلمات التي تظهر معها في السياق حسب طول السياق الذي يحدده المستفيد من خلال الكلمات السابقة أو اللاحقة، حيث يمكن للمستخدم اختيار طريقة إحصائية أو أكثر من ثمان طرق ممكنة في النظام لقياس قوة الترابط، إذ يُظهر النظام المعلومات التالية للمستفيد: الكلمة، تكرارها في المدونة، وتكرارها في السياق، وقيمة معامل الارتباط الإحصائي بين كلمتين أو أكثر، كما يسمح هذا النظام بترتيب النتائج تصاعدياً أو تنازلياً بالنقر على رأس العمود الذي يريد ترتيب النتائج بناء عليه، كما يمكن حفظ هذه النتائج في الحاسب الآلي الخاص بالمستخدم. والشكل الآتي يوضح صفحة التصاحب اللفظي في المدونة.

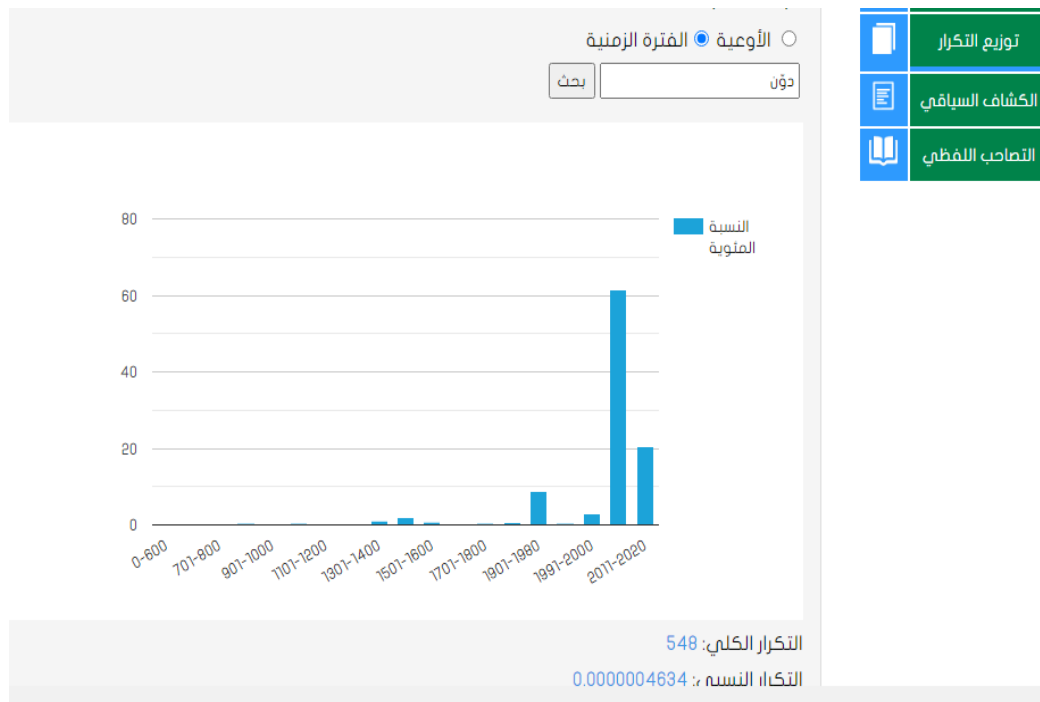
الشكل رقم (05): يوضح صفحة التصاحب اللفظي في المدونة اللغوية العربية لمدينة الملك عبد العزيز للعلوم والتقنية.



<https://corpus.kacst.edu.sa/collocation.jsp>

مثال تطبيقي على المدونة: -البحث عن كلمة -مدونة-

الشكل رقم (06) يوضح مثال تطبيقي لتوزيع تكرارات كلمة (دُون) على مدونة مدينة الملك عبد العزيز للعلوم والتقنية.



<https://corpus.kacst.edu.sa/collocation.jsp>

نلاحظ من خلال الشكل أنّ استعمال كلمة (دوّن) تكرّرت بكثرة في الفترة الزمنية ما بين 1991-2000، بنسبة 60%، في حين شهدت انخفاضا في الفترة ما بين 2011-2020 بنسبة 20%، ومنه فقد بلغ التكرار الكلي 548 مرّة، كما هو موضّح في الشكل.

2.4. المدوّنة اللّغويّة التاريخيّة للجامعة الأردنيّة: Historical Arabic Corpus

1.2.4. التعريف بالمدوّنة:

هذه المدوّنة متاحة عبر الإنترنت من خلال هذا الرابط: nlp.ju.edu.jo، وتهدف إلى خدمة علماء اللّغة و متعلّمي العربيّة، بحيث يمكنهم استكشاف وفهم الاستعمال اللّغوي وتطوّره، و التحقق من التغيّر الدلالي عبر العصور و المراحل الزمنية المختلفة للأدب العربي، ويبلغ حجم هذه المدوّنة 45 مليون هيكل كلمة من مختلف العصور التاريخيّة للأدب العربي، وذلك من خلال ما تحتويه المدوّنة من نصوص عربيّة قديمة تمتد لأكثر من ستّة عشر قرنا من الاستعمال اللّغوي؛ منذ العصر الجاهلي، وصدر الإسلام، والعصر العباسي الأوّل، والعصر العباسي الثاني، وعصور الولايات والممالك المتتاليّة، وحتى العصر الحديث، وكانت الشبكة العنكبوتيّة هي المصدر الأساس للحصول على هذه النصوص.

2.2.4. مميّزات المدوّنة:

تتيح هذه المدوّنة اللّغويّة إمكانيات استرجاع البيانات وفق قائمة تردد الكلمات و تركيب الجمل و العبارات، وعرض المتلازمات اللّغويّة لكلّ فترة زمنيّة أو نوع أدبي. وفيما يلي يوضح الشكل الصفحة الرئيسيّة للمدوّنة اللّغويّة التاريخيّة للجامعة الأردنيّة.

الشكل رقم (07): يوضح الصفحة الرئيسية للمدونة اللغوية التاريخية للجامعة الأردنية.

<https://corpus.kacst.edu.sa/collocation.jsp>

3.4. المدونة اللغوية العربية الدولية لمكتبة الإسكندرية: International Corpus of Arabic

1.3.4. التعريف بالمدونة:

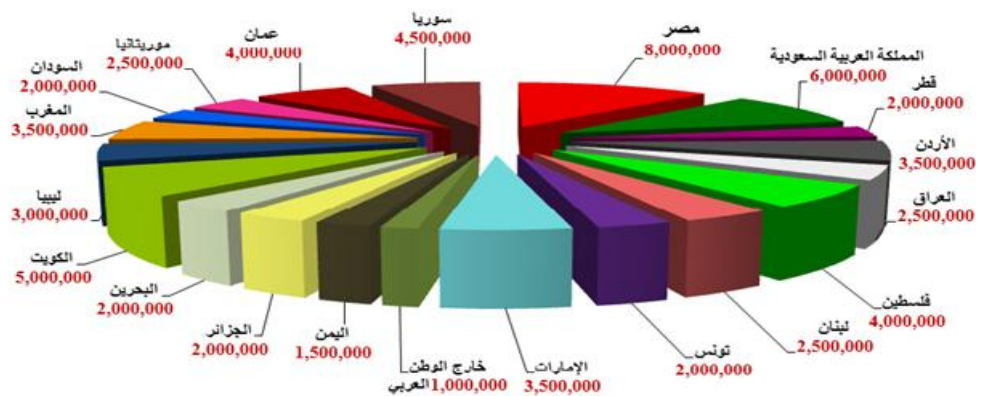
تمثل أحد المشروعات الثقافية التابعة لمكتبة الإسكندرية الهادفة لبناء مدونة لغوية للعربية المعاصرة، تحوي 100 مليون هيكل كلمة، محللة صرفيًا ونحويًا و دلاليًا، وممثلة لقطاع إقليمي كبير من الدول الناطقة باللغة العربية المعاصرة في أنحاء الوطن العربي¹⁵. وهي متاحة على الشبكة العنكبوتية تحت اسم "المدونة اللغوية العربية العالمية" من خلال هذا الرابط "www.bibalex.org/ica/ar/about.aspx"¹⁶، وفيما يلي شكل توضيحي للصفحة الرئيسية للمدونة بعد دخولنا للموقع:

الشكل رقم (08): يوضح الصفحة الرئيسية للمدونة اللغوية الدولية للمكتبة الإسكندرية.



الساعة: 18:54، بتاريخ: 2022/01/18، <http://www.bibalex.org/ica/ar/about.aspx>

الشكل رقم (09): يوضح تمثيل المدونة اللغوية الدولية للمكتبة الإسكندرية المنشورات في انحاء الوطن العربي.



الساعة: 18:54، بتاريخ: 2022/01/18، <http://www.bibalex.org/ica/ar/about.aspx>

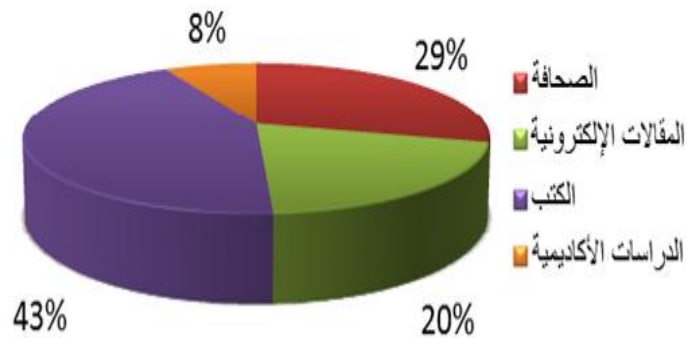
يعتمد تصميم هذه المدونة على البدء بمحصر المصادر المختلفة، وداخل كل مصدر يتم إدراج الفئات المميزة له، ويتم حفظ النصوص داخل هذه المدونة اللغوية بطريقة هرمية من خلال فهرسة النصوص وفقا لثلاثة من حقول الوصف، وهي: المصدر، والفئة، وتاريخ النشر.

توجد أربعة مصادر رئيسية تقوم عليها هذه المدونة اللغوية، هي: الصحافة (8%)، والمقالات الإلكترونية (20%)، والكتب (43%)، والدراسات الأكاديمية (8%)، وينقسم المصدر الخاص بالصحافة إلى ثلاثة مصادر فرعية، هي: الجرائد، والمجلات، والصحافة الإلكترونية، و يوجد إحدى عشرة فئة على مستوى المدونة اللغوية، هي: العلوم الإستراتيجية، والعلوم الإجتماعية، والرياضة، والدين، و الأدب، والعلوم الإنسانية، والعلوم الطبيعية، والعلوم التطبيقية، والفنون، و ثقافة، والسير الذاتية، والنصوص المتنوعة. وفيما يلي الشكل التوضيحي لنسبة هذه المصادر.¹⁷

الشكل رقم (10): يوضح النسبة المئوية التي تمثلها المصادر في المدونة اللغوية الدولية للمكتبة الإسكندرية.

تصميم المدونة اللغوية العربية العالمية

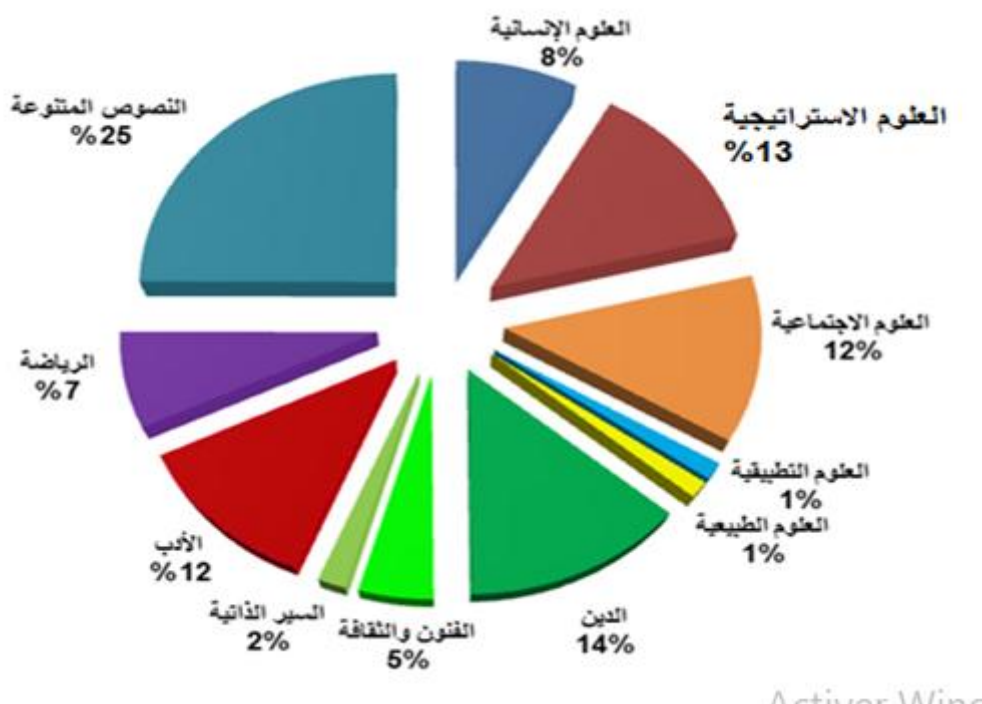
يوجد أربعة مصادر أساسية: الصحافة والمقالات الإلكترونية والكتب والدراسات الأكاديمية.



<http://www.bibalex.org/ica/ar/about.aspx>، الساعة: 19:15، بتاريخ 2022/01/18.

كما يوجد أربع و عشرون فئة فرعية، هي: السياسة، والقانون، والاقتصاد، والاجتماع، والدين الإسلامي، والدين المسيحي، و الأديان الأخرى، و الدين المقارن، والقصص، و الشعر، والنثر، و الدراسات اللغوية و الأدبية، و الطب، و الهندسة، والزراعة، والتكنولوجيا، وعلم الأحياء، وعلم الفيزياء، وعلم الفضاء، وعلم الجيولوجيا و البيئة، وعلم الكيمياء، وعلم النفس، وعلم الفلسفة، و التاريخ، ويوجد أربع فئات فرعية من فئة القصص الفرعية، هي: الروايات، والقصص القصيرة، وقصص الأطفال، والمسرحيات.

الشكل رقم (11): يمثل النسب المئوية لفروع المصادر التي تمثلها المدونة اللغوية الدولية للمكتبة الإسكندرية.



<http://www.bibalex.org/ica/ar/about.aspx>، الساعة: 19:15، بتاريخ 2022/01/18.

2.3.4. طريقة التحليل في المدونة:

يتم تحليل المدونة اللغوية بطريقة آلية مبنية على بعض الطرق الإحصائية، وبعض القواعد اللغوية بالاعتماد على أحد المحللات الصرفية الشهيرة، وهي مجرد بكوالتر Byckwalter stemmer، حيث يتيح التحليل عددا من إمكانات البحث، كالسوابق و اللواحق، وأقسام الكلمات، وساقها، وجذعها، وجذرها، ووزنها الصرفي، بالإضافة إلى نوع الكلمة من حيث نوع الجنس (ذكر/أنثى)، و العدد، و التعريف تبعا للسياقات المختلفة للكلمات داخل النصوص.¹⁸

4.4. المدونة اللغوية لمتعلمي اللغة العربية:

1.4.4. تعريفها:

هي مدونة متاحة على الشبكة من خلال الرابط: "www.arabiclearercorpus.com"¹⁹، يقوم عليها عبد الله الفيفي، الباحث بجامعة الإمام محمد بن سعود الإسلامية، بالمملكة العربية السعودية، ويتمثل الهدف من هذه المدونة توفير نصوص عربية مكتوبة و منطوقة حررها دارسو اللغة العربية في المملكة العربية

السعودية، على أن تكون هذه النصوص مفتوحة المصدر، كما وتحتوي هذه المدونة على مجموعة من النصوص والتسجيلات في موضوعين مختلفين: الأول سردي (رحلة خلال إحدى الإجازات)، و الثاني للمناقشة (الإهتمامات الدراسية)، وقد شارك في كتابة و تسجيل هذه المواد دارسو اللغة العربية في المملكة العربية السعودية خلال العامين: 2012 و 2013.²⁰

تشتمل المدونة على 282732 هيكل كلمة مع التكرار (386571 وحدة صرفية)، و 29627 كلمة دون تكرار، وهي عبارة عن 1585 مادة (مكتوبة أو مسجلة) انتجها 942 طالبا من 67 جنسية و 66 لغة أم مختلفة، تم تقسيمهم إلى مرحلتين: ما قبل الجامعة و المرحلة الجامعية، ويبلغ متوسط طول النصوص في هذه المدونة اللغوية 178 كلمة.

الشكل رقم (12): يوضح الصفحة الرئيسية للمدونة اللغوية لتعلمي اللغة العربية²¹



<https://www.arabiclearncorpus.com/home-ar>، الساعة: 19:38، بتاريخ: 2022/01/18.

2.1.4 الهدف من إنشاء المدونة²²:

- تزويد الباحثين بمجموعة من البيانات مفتوحة المصدر، للإستفادة منها في مجالات البحث اللغوي، مثل تعليم اللغة و تعلّمها، وعلم اللغة التطبيقي، وصناعة المعاجم.
- تحليل الأخطاء اللغوية.

- قياس التطور اللغوي لدى الطلاب، وتصميم المواد التعليمية.
- تأليف المعاجم الطلابية.

ملاحظة: يمكن البحث في المدونة من خلال المواقع الآتية:

الشكل رقم (13) يوضح المواقع التي يمكن البحث من خلالها في المدونة.²³

ابحث في نصوص المدونة من خلال المواقع التالية



<https://www.arabiclearnercorpus.com/home-ar>، الساعة: 19:42، بتاريخ: 2022/01/18.

5. دور المدونات اللغوية في بناء اقتصاد المعرفة:

يمكن استثمار المدونات اللغوية في بناء اقتصاد المعرفة، من خلال الاستفادة منها في مختلف أوجه مجالات المعرفة نحو الترجمة، و دراسة التوجهات الفكرية، و إسترجاع المعلومات، و صناعة المعاجم، وفي علم اللغة التطبيقي، وفيما يلي عرض لأوجه الاستفادة من المدونات اللغوية - العربية -²⁴:

1.5. في الترجمة:

تعدّ دراسة "ماثيو جيدير"، من أوائل الدراسات حول كيفية استخدام المدونات اللغوية العربية في الترجمة، حيث أشارت في هذه الدراسة أنّه يمكن منهجياً الجمع بين المقاربة اللغوية و المقاربة الإحصائية لضبط و رفع كفاءة معالجة المدونات اللغوية ثنائية اللغة بما يتّسق ومتطلبات الترجمة الآلية.

كما تعتبر دراسة المهنا في مجال الترجمة القائمة على المدونات اللغوية العربية، حيث إعتمدت على مدونة لغوية للأسماء و الصفات المركبة في المصطلحات الإنجليزية العلمية و التكنولوجية؛ وذلك في محاولة لنقلها إلى العربية. ومنه نقول أنّ المدونة اللغوية العربية تعتبر مكسب معرفي يسمح باستمرار التنمية المعرفية المستدامة.

2.5. في دراسة التوجهات الفكرية:

تعود أول دراسة عربية في تحليل التوجهات الفكرية، أو بالأحرى تحليل الخطاب النقدي، اعتماداً على المدونات اللغوية إلى دراسة الشمري و المحمود، حينما قام الباحثان بتحليل الخطاب النقدي لوعاء الأخبار اعتماداً على مدونة لغوية من قناتي العربية و الجزيرة حول أخبار الحرب على غزة في عام 2014، وبلغ عدد كلمات المدونة التي تم تحليلها 528641 كلمة موزعة على 1524 نصّاً إخبارياً.

3.5. في الإفادة من الدونات اللغوية في علم المعلومات:

تعتبر المدونات اللغوية أداة لا غنى عنها في اختبار و تقييم نظم استرجاع المعلومات، على النحو الذي يعكس الإستخدام الحقيقي لمدخلات هذه النظم، إضافة إلى إستخدامها في الكشف عن المعرفة في قواعد البيانات و التنقيب عن البيانات، و إقتباس/إستخراج المعلومات، و الكشف الدلالي الكامن، كما يمكن أن تستخدم كونها أداة للتحليل الكمي و النوعي في علم المعلومات، فعلى سبيل المثال، يمكن تجميع نصوص مجموعة من المقالات المتخصصة في علم المعلومات. كما أنّ المدونات اللغوية لا تقف عند حدود التحليل الكمي فحسب، بل تقوم أيضاً بتحليل نوعي يظهر البيئة اللغوية الداخلية، وكذلك البيئة المحيطة بالمادة المسترجعة.

4.5. الاستفادة من المدونات اللغوية في صناعة المكانز:

أتاحت بعض برمجيات معالجة المدونات اللغوية طرقاً للتحليل القواعدي للنصوص لتوليد و إنشاء المكانز، بفرضية أنّ المصطلحات المتشابهة أو المترادفة تظهر عادة في علاقات نظامية واحدة، حيث يتم تجميعها معاً وفقاً للسياق النحوي الذي وردت به، ومنه فإنّ المكانز تزداد بازدياد حجم المدونات اللغوية، فكلما ازداد حجم المدونة اللغوية، ازدادت القدرة على تعيين الكلمات المتشابهة و تمييزها عن غيرها، وقد أشرنا في عنصر سبق إلى أهمّ المكانز العربية.

5.5. الإفادة من المدونات اللغوية في صناعة المعاجم:

أحدثت المدونة اللغوية ثورة في صناعة المعاجم حتى أضحت أداة رئيسية لصانعيها، لاسيما الضخمة منها، حيث تسهم في رصد الكلمات الجديدة التي تدخل في اللغة و تحدّد وقت دخولها، كما ويستطيع صانعو المعاجم بمساعدة المدونات اللغوية البحث في ملايين الكلمات و الجمل، وسياقاتها المختلفة، وإسترجاع كافة الأمثلة التي وردت من أجل تحري استخداماتها، وتحديد التلازم اللغوي، والتعبيرات الإصطلاحية بسهولة.

6.5. الإفادة من المدونات اللغوية في تدريس و تعلّم اللغات:

أضحت المدونات اللغوية أداة رئيسة لدى الكثير من معلمي و دارسي اللغات، حيث يمكن أن تفيد كثيرا في مجال تدريس و تعلّم اللغات من جهة الباحثين والطلاب و المعلمين على حدّ سواء.

7.5. الإفادة من المدونات في علم اللغة الاجتماعي:

تُمثّل المدونات اللغوية أحيانا وفقا للسن أو النوع، أو المستوى التعليمي، أو الاجتماعي، أو الإقتصادي أو المكان الجغرافي... أو غيرها من العوامل و المتغيرات الاجتماعية التي يتمّ بها تمثيل النصوص في المدونات اللغوية، فعلى سبيل المثال قد يتم الإعتماد على مدونة لغوية لمقالات أو أخبار مجموعة من الصحف العربية المنشورة ببلدان عربية مختلفة، ومن ثمّ يمكن التحقّق من التنوع اللغوي في لغة الإعلام لهذه البلدان في ضوء المطابقة النحوية، أو الإملاء، و إختلافاتها المتأثرة بالعوامل الجغرافية.

نستطيع القول أنّ المدونات اللغوية عموما والمدونات اللغوية العربية على وجه الخصوص، يمكن إستثمارها في مجال إقتصاد المعرفة من خلال إستثمارها في علم اللغة التطبيقي (النحو، والدلالة، وعلم اللغة الاجتماعي، وصناعة المعاجم، والترجمة، وتحليل التوجهات الفكرية)، ودراسات الاستخدام في إسترجاع المعلومات، وبناء المكنز.

الجدول رقم (01) ملخص المدونات اللغوية العربية الشهيرة.

اسم المدونة	مكانها - صاحبها	نوعها	حجمها	غرضها	مصادر جمعها
المدونة اللغوية العربية لمدينة الملك عبد العزیز للعلوم والتقنية	المملكة السعودية العربية	رقمية	182515633 كلمة مصنفة على 80 نطاقاً، و 481 موضوعاً	- تطوير تطبيقات الدراسات اللغوية بمستوياتها المتنوعة. - معالجة اللغة الطبيعية	- المخطوطات المحققة. - الكتب - الصحف - المجلات - المناهج الدراسية. - الرسائل العلمية - المواقع الإلكترونية - الدوريات المحكمة. - المطبوعات الرسمية. - وكالات الأنباء.
المدونة اللغوية للجامعة الأردنية	الأردن	رقمية	45 مليون هيكل كلمة	- خدمة علماء اللغة ومتعلمي العربية. - فهم الاستعمال اللغوي وتطوره، والتحقق من التغير الدلالي عبر العصور.	الشبكة العنكبوتية
المدونة اللغوية العربية الدولية لمكتبة الإسكندرية	القاهرة، مصر	رقمية	100 مليون هيكل كلمة	- التحليل الصرفي والنحوي والدلالي	- الصحافة (29%) - المقالات الإلكترونية (20%) - الكتب (43%) - الدراسات الأكاديمية (8%)
مدونة عربي كورنيس	جامعة بريجهام الأمريكية	رقمية	173600000 هيكل كلمة	- توفير إمكانات البحث.	- الصحف
المدونة اللغوية لتعليمي اللغة العربية	- جامعة الإمام محمد بن سعود الإسلامية. - عبد الله الفيفي	رقمية ورقية	282732 هيكل كلمة تكرار 386571 حدة صرفية.	- توفير نصوص عربية مكتوبة و منطوقة. - تزويد الباحثين بمجموعة من البيانات مفتوحة المصدر للاستفادة منها في مجالات البحث اللغوي.	- أصول الأوراق المتاحة يدوياً.
مدونة قرآني	الأزهر الشريف	رقمية	1200 مصطلح	- البحث عن المفاهيم الواردة بآيات القرآن الكريم، والكشف عن سياقه (تكشيف النصوص).	مصحف التجويد للدكتور محمد حبش.

المصدر: من إعداد الباحثين

خاتمة:

بعد العرض التفصيلي لأهمّ المدوّنات اللّغويّة العربيّة الشهيرة، والوقوف على أهمّ خصائصها، تمّ التوصل إلى النتائج الآتية:

- المدوّنات اللّغويّة العربيّة أداة لا غنى عنها في إختبار وتقييم نظم إسترجاع المعلومات.
- يمكن أن تستخدم المدوّنات اللّغويّة العربيّة كأداة للتحليل الكميّ والتّوعي في علم المعلومات.
- تعتمد المدوّنات اللّغويّة العربيّة في عملها على قوّة الحاسبات الآليّة في الاختزان، وقدرات البرمجيات المختلفة في التكشيف، والتحليل والاسترجاع.
- المدوّنات اللّغويّة لا تقف عند حدود التحليل الكميّ فحسب، بل تقوم أيضا بتحليل نوعيّ يظهر البيئة اللّغويّة الداخليّة، وكذلك البيئة المحيطة بالمادة المسترجعة.
- يمكن للمدوّنات اللّغويّة العربيّة أن تحدّد الفترة الزمنيّة للكلمة، والكشف عن سياقها، والألفاظ المصاحبة لها، كما تقدّم إحصاءات حول تكرارها.
- تعكس المدوّنات اللّغويّة العربيّة الإستعمال الحقيقي أو الواقعي للغة في شكل مقروء آليّا.
- المدوّنات اللّغويّة العربيّة مفتاح للمشروعات العلميّة والعملية.
- تحقّق المدوّنات اللّغويّة العربيّة تراكما معرفيّا ثريا.
- تمكّن المدوّنات اللّغويّة العربيّة من تحقيق الوجود السامي للغة العربيّة في جميع مجالات الحياة عموما، وفي اقتصاد المعرفة خصوصا.
- يمكن للمدوّنات اللّغويّة أن تجيب عن الأسئلة التالية: ما أكثر العبارات تردّدا؟ ما أوجه الإختلاف بين النصوص المكتوبة والنصوص المنطوقة؟ ما الأفعال، أو الأسماء أو الحروف التي يستخدمها أهل اللّغة؟

- يمكن الاستفادة من المدونات اللغوية العربية في صناعة المعاجم، وبناء المكانز، وتدريس اللغات، ودراسة التوجهات الفكرية (الإيديولوجية)، والترجمة.

الهوامش

¹ - Khosrow-pour, in Encyclopedia of Information Science and Technology, third edition, 2015, p4151.

² - رياض خليل ابراهيم، علم لغة المدونات، والترجمة، مجلة جامعة بابل الإنسانية، المجلد 26، العدد 8، 2018، ص 150.

³ - احمد عبد الغني محمد، الاتجاه المعتمد على المدونات اللغوية في تعليم اللغة العربية للناطقين بغيرها، مجلة كلية التربية للبنات، جامعة بغداد، 2020، ص 3.

⁴ - أيمن الذكوروي، المدونات اللغوية و دورها في معالجة النصوص العربية، ط 1، دار وجوه للنشر و التوزيع، الرياض، السعودية، 1439-2018، ص 27.

* في الغالب تكون المدونة محدّدة الحجم و دائما يتمّ تحديد حجم المدونة في طور تصميمها، وقبل الشروع في جمعها، ولكن هذا ليس مطلقاً حيث يوجد نوع من المدونات - يطلق على اسم "monitor corpus" - تكون غير محدّدة الحجم وفي تزايد مستمر، ولكن باستثناء هذا النوع من المدونات، فأى مدونة تكون محدّدة الحجم.

⁵ - ربيع بن المر الدهلي، رحمة بنت ناصر الراشدية، دور جامعة نزوى في نشر مفهوم إقتصاد المعرفة لدى الطلبة من وجهة نظر أعضاء هيئة التدريس، مجلة البحوث التربوية و النفسية، المجلد 19، العدد 72، جامعة بغداد، 2022، ص 207.

⁶ - محمد ابراهيم علي، إقتصاد المعرفة و المحاسبة الدولية في تحقيق التنمية المستدامة، مجلة العلوم الإقتصادية و الإدارية، المجلد 21، العدد 81، كلية بغداد، 2015، ص 402.

⁷ - عبد العزيز المهوي، رسائل الدراسات العليا في الجامعات العربية في خدمة اللغة العربية حاسوبياً، مجلة مجمع اللغة العربية الأردني، العدد 100، ص 150.

⁸ - أيمن الذكوروي، المدونات اللغوية ودورها في معالجة النصوص العربية، ط 1، دار وجوه للنشر و التوزيع، الرياض، السعودية، 1439هـ-2018، ص 92.

⁹ - corpus.kacst.edu.sa/index.jsp

¹⁰ - corpus.kacst.edu.sa/index.jsp

¹¹ - corpus.kacst.edu.sa/index.jsp

¹² - corpus.kacst.edu.sa/index.jsp

¹³ - corpus.kacst.edu.sa/index.jsp

¹⁴ - corpus.kacst.edu.sa/index.jsp

¹⁵ - <http://www.bibalex.org/ica/ar/about.aspx>، الساعة: 16.36، بتاريخ: 2022/01/17

¹⁶ - أيمن الذكوروي، مرجع سابق، ص 101.

¹⁷ - <http://www.bibalex.org/ica/ar/about.aspx>، الساعة: 19:15، بتاريخ: 2022/01/18.

¹⁸ - <http://www.bibalex.org/ica/ar/about.aspx>، الساعة: 19:20، بتاريخ: 2022/01/18.

¹⁹ - أيمن الذكوروي، مرجع سابق، ص 104.

²⁰ - <https://www.arabiclearnercorpus.com/home-ar>، الساعة: 19:42، بتاريخ: 2022/01/18.

²¹ - <https://www.arabiclearnercorpus.com/home-ar>، الساعة: 19:38، بتاريخ: 2022/01/18.

²² - <https://www.arabiclearnercorpus.com/home-ar>، الساعة: 19:42، بتاريخ: 2022/01/18.

²³ - <https://www.arabiclearnercorpus.com/home-ar>، الساعة: 19:42، بتاريخ: 2022/01/18.

²⁴ - ينظر، أيمن الدكروني، مرجع سابق، ص 141-142-143-144-145-146-147.

قائمة المراجع:

الكتب:

- أيمن الدكروني، المدونات اللغوية و دورها في معالجة النصوص العربية، ط1، دار وجوه للنشر و التوزيع، الرياض، السعودية، 2018-1439.
- Khosrow-pour, in Encyclopedia of Information Science and Technology, third edition, 2015

المقالات:

- أحمد عبد الغني محمد، الاتجاه المعتمد على المدونات اللغوية في تعليم اللغة العربية للناطقين بغيرها، مجلة كلية التربية للبنات، جامعة بغداد، 2020.
- ربيع بن المر الذهلي، رحمة بنت ناصر الراشدية، دور جامعة نزوى في نشر مفهوم إقتصاد المعرفة لدى الطلبة من وجهة نظر أعضاء هيئة التدريس، مجلة البحوث التربوية و النفسية، المجلد 19، العدد 72، جامعة بغداد، 2022.
- رياض خليل ابراهيم، علم لغة المدونات، والترجمة، مجلة جامعة بابل الإنسانية، المجلد 26، العدد 8، 2018.
- عبد العزيز المهيوبي، رسائل الدراسات العليا في الجامعات العربية في خدمة اللغة العربية حاسوبياً، مجلة مجمع اللغة العربية الأردني، العدد 100، د.س.
- محمد ابراهيم علي، إقتصاد المعرفة و المحاسبة الدولية في تحقيق التنمية المستدامة، مجلة العلوم الإقتصادية و الإدارية، المجلد 21، العدد 81، كلية بغداد، 2015.

المواقع الإلكترونية:

- corpus.kacst.edu.sa/index.jsp
- <https://corpus.kacst.edu.sa/index.jsp/> <https://corpus.kacst.edu.sa/tools.jsp>
- https://corpus.kacst.edu.sa/stem_search.js
- https://corpus.kacst.edu.sa/word_frequency.jsp
- <http://www.bibalex.org/ica/ar/about.aspx>
- <https://www.arabiclearnercorpus.com/home-ar>