

حوسبة المعاجم العربية (ومحرك بحث معجمي)

<https://ontology.birzeit.edu/about-ar>

تمت حوسبة 150 معجم عربي يدوياً وربطها وتوحيدها في قاعدة بيانات واحدة لتصبح أضخم قاعدة بيانات معجمية في تاريخ اللغة العربية. تحتوي قاعدة البيانات على أنواع كثيرة من المصادر اللغوية: كمعاجم لغوية قديمة ومعاصرة، وقواميس ثنائية وثلاثية اللغة، وقواعد بيانات تصريفية واشتقاقية، ومسارد مصطلحات، ومكانز مترادفات، ومعاجم فروق لغوية، وغيرها. وهي بذلك تغطي معظم الحقول، مثل العلوم، والهندسة، والصحة، والفلسفة، والإنسانيات، والفنون، والقانون.

- المعاجم متاحة للعامة عبر محرك بحث معجمي يتيح استرجاع تعريفات، ومترادفات، وفروق لغوية، وترجمات متخصصة. بالإضافة إلى إتاحتها لمطوري البرمجيات عبر واجهة برمجية (API).
- محرك البحث المعجمي هو الأول من نوعه عالمياً، فمثلاً، لا يوجد محرك بحث للإنجليزية يحتوي هذا الكم من المعاجم في قاعدة بيانات واحدة.
- استغرق العمل على هذا المشروع 10 سنوات، وهو عمل بحثي غير ربحي، ولم يُمول من أية جهة.
- تم التواصل مع أصحاب المعاجم وأخذ إذنهم احتراماً لحقوق الملكية، كما ونُظهُرُ معلومات المعجم إلى جانب كل تعريف (المؤلف، الناشر، ...) لغرض الترويج للمعجم.
- فاز المشروع بجائزة محمد بن راشد للغة العربية، وغطته العديد من الفضائيات والإذاعات والصحف.
- نُشِرَت عدة أوراق علمية وصفت المشروع من الناحية البحثية (حوسبة اللغة).

أوراق علمية ذات علاقة بحوسبة المعاجم للمزيد يمكن الضغط على الرابط بلون (تركواز)

Mustafa Jarrar, Hamzeh Amayreh: **An Arabic-Multilingual Database with a Lexicographic Search Engine**. The 24th International Conference on Applications of Natural Language to Information Systems (NLDB 2019). Pages(234-246). LNCS 11608, Springer. 2019
* وصف لمنهجية حوسبة المعاجم، محرك البحث، وإحصاءات مفصلة لقاعدة البيانات المعجمية

Mustafa Jarrar, Hamzeh Amayreh, John P. McCrae: **Representing Arabic Lexicons in Lemon - a Preliminary Study**. The 2nd Conference on Language, Data and Knowledge (LDK 2019). Pages(29-33). CEUR, Volume 2402. ISSN:1613-0073. Leipzig, Germany. 2019
الورقة تقترح تمثيل البيانات المعجمية بمقاييس (W3C Lemon Model) وربطها بالشبكة المعجمية العالية

Diana Alhafi, Anton Deik, Mustafa Jarrar: **Usability Evaluation of Lexicographic e-Services**. The 16th IEEE/ACS International Conference on Computer Systems and Applications (AICCSA). Pages(1-7). IEEE. Abu Dhabi, UAE. 2019
تقييم مدى سهولة استخدام محرك البحث المعجمي ومقارنته بخدمات غوغل المعجمية

Mustafa Jarrar: **Search Engine for Arabic Lexicons**. The 5th Conference on Translation and the Problematics of Cross-cultural Understanding. The Forum for Arab and International Relations. Doha, Qatar. December, 2018
ورقة (بالعربية) حول حوسبة المعاجم

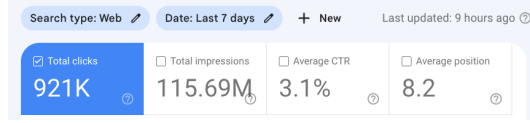
Mustafa Jarrar: **Digitization of Arabic Lexicons**. Arabic Language Status Report. UAE Ministry of Culture and Youth. Pages 214-2017. Dec 2020
ورقة (بالعربية) تصف مشروع حوسبة المعاجم، ظهرت في تقرير حالة اللغة العربية

Hamzeh Amayreh, Mohammad Dwaikat, and Mustafa Jarrar: **Lexicon Digitization -A Framework for Structuring, Normalizing and Cleaning Lexical Entries**. Technical Report. Birzeit University. 2019
تقرير علمي يصف الأدوات والبرمجيات التي تم تطويرها لحوسبة وربط المعاجم العربية

Mustafa Jarrar, Fadi Zaraket, Rami Asia, Hamzeh Amayreh: **Diacritic-Based Matching of Arabic Words**. ACM Asian and Low-Resource Language Information Processing. Volume 18, No 2, Pages(10:1-10:21), ACM, ISSN:2375-4699. December, 2018
ورقة تصف خوارزمية لمقارنة الكلمات العربية (وربط المدخلات العربية)

مدى استخدام محرك البحث المعجمي

الرسم التالي (من Google Webmaster) يوضح الاستخدام آخر سبعة أيام: حوالي 921 ألف عملية بحث.



في الإعلام (فضائيات وصحف وإذاعات ومحاضرات) حول حوسبة المعاجم العربية

Birzeit Won Bin-Rashid's "Arabic Language Award" for its Lexicographic Search Engine (بالعربية)



Birzeit University was awarded Mohammed Bin Rashid's Award for the Arabic Language 2019 for its innovative Lexicographic Search Engine -the first of its kind worldwide. The engine is developed on top of the 150 lexicons digitized and unified over ten years by Prof Jarrar and his students.



تقرير قناة العربية



Keynote at ARDD: The Arabic Lexicographic Database and Ontology (بالعربية)

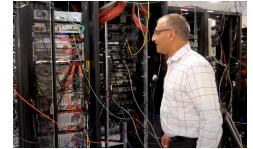


Keynote by Prof. Jarrar presenting eight years of research to enrich Arabic content on the Internet through a lexicographic search engine and the Arabic ontology - to enable NLP applications to better understand Arabic semantics. The keynote was organized by the ARDD organization, and was attended by about 70 scientists in Amman.



تقرير قناة بالعربي

(حوسبة الدلالات)



An Online Arabic Dictionary Makes Its Debut (بالعربية)



A professor at Birzeit University has built an online tool that is both a comprehensive dictionary of Arabic and a system that will enable the creation of new Arabic-language software. Prof. Mustafa Jarrar has worked for eight years to create this new tool, not only as a searchable lexicon but also as a logical system—an ontology—that recognizes the unique characteristics of the Arabic language.



مقابلة قناة رؤيا



Birzeit announces Linguistic search engine and the Arabic Ontology (بالعربية)



A linguistic search engine with 150 lexicons and a linguistic ontology for Arabic was launched on September 25, 2018. The ceremony was attended by the University President, Deputy Undersecretary of the Ministry of Education, President of the Palestine Academy for Science and Technology, Secretary General of the National Commission for Education, Culture and Science, and UN-ESCWA's Chief of Innovation.



تقرير تلفزيون فلسطين



مقابلة قناة فلسطين



An Arabic Greeting to Birzeit University

(بالعربية)



An article by Abu baker Saadallah, an Algerian Scientist, describing the pioneering role of Birzeit University in the region since 1924, and focusing on the Arabic Ontology and the digitization of the Arabic lexicons. ومهما يكن من أمر، فاللغة العربية لن تتوقف مسيرتها نحو الأفضل ما دام في البلاد العربية نساء من أمثال نبيهة ناصر، ورجال من أمثال مصطفى جرار، ومؤسسات كجامعة بيرزيت من وراء القصد. فتحية عربية لهم جميعا... بمناسبة اليوم العالمي للغة

العربية!



more→

MoU with ALECSO to digitize 50 dictionaries

(بالعربية)



Birzeit University and the Arab League Educational, Cultural and Scientific Organization (ALECSO) signed a memorandum of understanding (MoU) to digitize 50 new dictionaries and to establish a linguistic observatory, in addition to cooperation on topics related to computational linguistics and language engineering.



more→

Interview about digitization and engineering of dictionaries

(بالعربية)



An interview with Dr Mustafa Jarrar hosted at the Palestine Today TV satellite channel. During which, Dr Jarrar explained the process of digitization and integration of Arabic dictionaries. He also answered questions about the Arabic Ontology, and Google Research Award that he received for fostering R&D activities related to publishing a large-scale lexicon database for Arabic.



more→

Researcher wins Google Research Award

(بالعربية)



Birzeit University computer science professor Mustafa Jarrar has won Google's Faculty Research Award in the field of natural language processing. Jarrar received a grant of \$50,000 to help further develop and support his research and enable him to launch an online portal comprised of a large number of Arabic lexicons



more→

مقابلة فضائية القدس



مقابلة القدس التعليمية



إطلاق محرك البحث



حوسبة المعاجم العربية

دراسة حالة



بقلم: د. مصطفى جرّار
جامعة بيرزيت، فلسطين

مقدّمة

تعرض هذه الورقة ملخصاً لتجربة بناء أضخم قاعدة بيانات لغويّة للغة العربيّة، وكذلك محرّك بحث للمعاجم العربيّة يحتوي على 150 معجمًا تمّت حوسبتها في جامعة بيرزيت بفلسطين.¹⁴² ومحرّك البحث هذا متاح للعامة ولجمهور الطلبة والباحثين والمترجمين ومتعلّمي اللغة وغيرهم، وهو يتيح للمستخدم البحث عن الكلمة في 150 معجمًا عربيًا ومتعدّد اللغات واسترجاع تعريفاتها ومرادفاتها وكذلك ترجماتها المتخصّصة. وهذه الترجمات تعتمد على المعاجم المدقّقة والمنقّحة، وهي أدقّ بكثير من جميع المترجمات الآليّة المتاحة (مثل جوجل للترجمة)، وتجدر الإشارة إلى أنّ محرّك البحث هو الأوّل عالميًا، حيث لا يوجد محرّك بحث لمعاجم أيّ لغة أخرى، حتّى الإنجليزيّة. ويمكن لمطوّري البرمجيات كذلك استعمال محرّك البحث عبر واجهة خاصّة (API) للوصول إلى الترجمات والمترادفات واستخدامها داخل برمجياتهم وتطبيقاتهم. وقد تمّت حوسبة المعاجم بطريقة يدويّة على مدى تسع سنوات، وبعد ذلك قمنا بتوحيد هذه المعاجم في قاعدة بيانات واحدة تشتمل على معاجم لغويّة تقليديّة وحديثة، ومسارد، ومكانز، ومعاجم ثنائيّة وثلاثيّة اللغة، وقواعد بيانات تصريفية واشتقاقية. وتعدّ الأنطولوجيا العربيّة أهمّ المصادر اللغويّة في محرّك البحث، وهي شجرة أو تصنيف لمعاني الكلمات العربيّة وتمثّل لها بلغة المنطق، بحيث يستطيع الحاسوب فهم المعاني ومعالجتها. وقد تمّ بناء هذه الشجرة كمشروع بحثي منفصل في جامعة بيرزيت،¹⁴³⁻¹⁴⁴ ونعمل حاليًا على ربط جميع مدخلات المعاجم

وصولًا إلى شبكة لغويّة محوسبة للغة العربيّة تُسمّى Big Linguistic Data Graph، ومن ثمّ ربط هذه الشبكة باللغات الأخرى.¹⁴⁵

وبأني هذا العمل ضمن مشروع بحثي غير ربحي وطويل الأمد لخدمة اللغة العربيّة، ولإغناء الإنترنت بمحتوى عربيّ نوعي، إذ سيتمكّن المستخدمون من إيجاد المصطلح العربيّ الذي يناسب متطلّباتهم وترجمات متعدّدة، خاصّة وأنّ قاعدة البيانات تحتوي على عدد ضخم من معاجم المصطلحات المعاصرة وفي شتى العلوم والمجالات العلميّة والهندسيّة والتجاريّة والأدبيّة وغيرها. إضافةً إلى ذلك، فإنّ محرّك البحث متاح للعامة مجانًا من خلال موقع جامعة بيرزيت الإلكترونيّ (<https://ontology.birzeit.edu>) الذي يقدّم تعريفًا بالمحرّك وبالمشاريع الأخرى ذات العلاقة على هذا الرابط (<https://ontology.birzeit.edu/about-ar>)، بالإضافة إلى روابط للأوراق العلميّة التي بُني محرّك البحث على أساسها. وتجدر الإشارة إلى أنّ محرّك البحث تمّ تطويره وفق المعايير والمقاييس الصادرة عن منظّمة الشبكة العالميّة (W3C)، وخاصّة المعايير المتعلّقة بأسلوب نشر البيانات على الإنترنت، وكذلك المعايير المتعلّقة بتمثيل وتبادل البيانات اللغويّة.

وفيما يلي يقدّم الشكل التالي مثالاً على طريقة استرجاع ترجمات كلمة «attribute» من عدّة معاجم ومن الأنطولوجيا، ويستطيع الباحث هنا التحكّم فيما إذا كان يرغب في استرجاع ترجمات، و/أو مترادفات، و/أو تعريفات فقط. وتظهر نتائج المعاجم في الجهة اليسرى من الشكل، ونتائج الأنطولوجيا في الجهة اليمنى:



attribute


 Translations Synonyms Definitions

[Ontology](#) [Dictionaries](#) [Morphology](#)
[About](#) [License](#)

Like Share 34 results (0.50 secs)

attribute pridecate محمول
Propositions منطقياً - ما يقال على موضوع ومنه القضايا الحملية
attributives

Philosophy Lexicon ©

attribute pridecate محمول
قال ابن سينا: "المحمول هو المحكوم به أنه موجود أو غير موجود لشيء آخر"

Philosophy Lexicon ©

attribute pridecate محمول
ميتافيزيقياً: ويسمى صفة، وهي خاصة ذاتية للجوهر، ويطلق بوجه خاص على صفات الباري. ومشكلة الإسلامي، وحولها دار معظم نشاط المدارس الكلامية.

Philosophy Lexicon ©

attribute صفة
The Unified Dictionary of Social Sciences Terms ©

attribute صفة مُمَيَّزَة | سَمَة
صفة تميز ملفاً أو سجلاً أو حقلاً وتبين نوعه أو كيفية التعامل معه، أو تبين أي معلومات عن البيانات المخزونة فيه.

Computers Lexicon ©

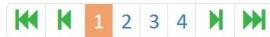
attribute صفة | خاصية
Statistical Terms ©

inspection by attribute الفحص بالصفات
Statistical Terms ©

sampling for attribute المعاينة للخاصية | المعاينة للصفة
Statistical Terms ©

quality attribute مِغيار الجودة
إحدى الخواص التي تساهم في جودة المادة.
The Unified Dictionary of Nutrition Technologies Terms ©

attribute samples عَيِّنات وصفية
The Unified Dictionary for terminologies of General and Nuclear Physics Terms ©



ONTOLOGY

attribute - 1 results (0.28 secs)

attribute | صِفَةٌ | خَصْلَةٌ | مُسَنَدٌ | حَالَةٌ | نَعَتْ | مَرِيَّةٌ | مَبْرَةٌ | شَكْلٌ | هَيْئَةٌ | سِمَةٌ

An abstract that describes qualities of an entity without using units of measure

مجرد يصف ويُعبّر عن قيمة خاصية لشيء ما دون استعمال وحدة قياس
example: اللون البرتقالي هو صفة لخاصية اللون في البرتقال
293263 TypeOf : {abstract}

Physical Attribute صِفَةٌ مَادِّيَّةٌ
An attribute that describes the value of a physical quality

صفة للتعبير عن قيمة خاصية مادية
الأخضر هي صفة مادية لخاصية لون عيون بعض الأشخاص
293265 TypeOf : {attribute}

Abstract Attribute صِفَةٌ مَجْرَدَةٌ | صِفَةٌ إِعْتِبَارِيَّةٌ
Attribute that describes an abstract quality of things

صفة تعبّر عن قيمة خاصية مجردة لشيء ما
كريم هي صفة مجردة لخاصية العطاء عند الإنسان
293264 TypeOf : {attribute}

الرَّسْمُ التَّامُّ
ما يتركب من الجنس القريب والخاصة كتعريف الإنسان بالحيوان الضاحك
41816 TypeOf : {attribute} ©Al-Jirjani Definitions

الرَّسْمُ النَّاقِصُ
ما يكون بالخاصة وحدها أو بها وبالجنس البعيد كتعريف الإنسان بالضحك أو بالجسم الضاحك أو بعرضيات تختص جملتها بحقيقة واحدة كقولنا في تعريف الانسان أنه ماشي على قدميه عريض الأظفار للمزيد...
41817 TypeOf : {attribute} ©Al-Jirjani Definitions

الاستخدام الورقي، بينما يحتاج الاستخدام الحاسوبي إلى تراكيب بنويّة مختلفة تسهّل عمليّات الإضافة والاسترجاع والبحث.

وتشتمل قاعدة البيانات اللغويّة الحاليّة على أنواع مختلفة من المعاجم، مثل المعاجم اللغويّة التقليديّة القديمة والجديدة، وكذلك المسارد (glossaries) التي تشرح المصطلحات، والمكانز (thesauri) التي تحتوي على مترادفات، ومعاجم ثنائيّة وثلاثيّة اللغة، ومعاجم الفروق اللغويّة، وقواعد بيانات تصنيفيّة واشتقاقية. كما تغطّي قاعدة البيانات شتى المجالات مثل العلوم الطبيعيّة، والهندسة، والطب، والاقتصاد، والأدب، والإنسانيّات، والفلسفة،

حوسبة المعاجم

قمنا بحوسبة حوالي مائة وخمسين معجمًا عربيًا ومتعدّد اللغات، واضطررنا لطباعة غالبيتها يدويًا، خاصّةً وأنّه لا توجد نسخ إلكترونيّة للغالبية المعاجم حتّى في حوزة مؤلّفيها. وفي المراحل التالية، عملنا على تطوير برمجيات وخوارزميات لإعادة تشكيل بنويّة (re-structuring) لمدخلات المعاجم،¹⁴⁶ بحيث يتمّ الفصل بين المدخلة المعجميّة والتعريف والمشتقات والتصريفات والترجمات والمترادفات، انتهاءً بحفظها في قاعدة بيانات واحدة.¹⁴⁷ وجاءت الخطوة الأخيرة ضروريّة لكون صناعة المعاجم التقليديّة تركّز في تصميمها على

1- عدم توفرّ مصادر لغوية محوسبة

بالرغم من وجود كمّ كبير من المعاجم العربيّة المطبوعة ورقياً، فإنّ عدد المصادر اللغويّة المحوسبة والمتاحة للبحث العلميّ ولمطوّري البرمجيات ضئيل، ممّا يحّد من دعم اللغة العربيّة في معظم التطبيقات الحاسوبية. وعلى الرغم من أنّ سوق التقنيّات في العالم العربيّ كبير، فإنّ مطوّري البرمجيات يواجهون تحديّات كبيرة في إيجاد مصادر لغوية يمكن استعمالها بسهولة وبتكلفة معقولة.

2- ضعف الصناعة المعجمية وتوانيتها عن مواكبة النظريّات اللسانية الحديثة

ثمّة نقص في أنواع جديدة من المعاجم، مثل: مكانز المترادفات، ومعاجم الفروق الدلالية، والشبكات الدلالية، ومعاجم السلوك التصريفيّ للأفعال، ومعاجم المصطلحات والترجمات الحديثة والكلمات العربيّة المستجدة. وقد أصبحت الحاجة إلى مثل هذه الأنواع من المعاجم، والتي لم تكن موجودة أو منتشرة قديماً، ملحة مع تطوّر احتياجات المجتمع، ووسائل تعليم اللغة، وظهور برمجيات ذكية أصبحت اللغة ركنًا أساسياً فيها.

3- عدم وجود منهجيات موحّدة ومعيارية في الصياغة المعجمية العربيّة

فمثلاً، لا يوجد اتّفاق على كيفية صياغة ووسم المدخلات المعجمية، ويظهر ذلك جليّاً في التباين الحاصل بين صياغة المعاجم التي صدرت في السنوات الأخيرة.

4- عدم توفرّ قواعد بيانات لغوية شاملة

فعلى سبيل المثال، لا توجد قائمة شاملة وموحّدة بالجزور، أو بالأوزان الصرفية والاشتقاقية، أو بفروع الكلمات (lemmas)، أو بالاشتقاقات، أو غيرها. ويشكّل هذا النقص تحديّاً لكلّ من يريد المساهمة في تأليف معجم، أو إنشاء قاعدة بيانات لغوية، أو بناء تطبيق حاسوبيّ.

5- عدم توفرّ ميزاتٍ للبحث العلميّ الرصين والموجّه نحو حوسبة اللغة العربيّة لدعمها في التطبيقات الحاسوبية.

تتركز أبحاثنا في جامعة بيرزيت حالياً في ثلاثة مسارات:

الأول: إغناء المحرّك المعجميّ بمزيد من التعريفات والشروحات اللازمة، وتوسيع الأنطولوجيا العربيّة بإضافة المزيد من المفاهيم في المستويات الدنيا من الشجرة الدلالية.

الثاني: ربط مدخلات المعاجم العربيّة مع شبكة المفردات الإنجليزيّة (Wordnet)، وهي شبكة مفردات تمّ تطويرها في جامعة برنستون وربطها بغالبية اللغات في العالم،¹⁵⁶ وبالتالي، فإنّ ربطها بالمدخلات المعجمية العربيّة يؤدي تلقائياً إلى ربط المدخلات العربيّة بسائر اللغات العالمية الأخرى.

الثالث: وهو المشروع المستقبليّ الذي شرعنا بأولى خطواته، ويقوم على ربط المعاجم وجميع البيانات اللغوية المتوفرة لدينا ببعضها، وصولاً إلى بناء شبكة بيانات لغوية شاملة من حيث الكمّ ومن حيث المستويات اللغوية الدلالية والتصريفية والاشتقاقية، والمقابل في اللغات الأخرى. وبدوره، فإنّ هذا المسار يتضمّن ثلاثة مستويات مترابطة:

- تصريفياً. نعمل على وصف كلّ مدخلة معجمية بجذعها (lemma)، وبالتالي تصبح كلّ مدخلات المعاجم مترابطة تصريفياً، بغضّ النظر عن السوابق واللواحق في المدخلات، أو الصيغ المختلفة من نفس المدخلة في المعاجم الأخرى.
- اشتقاقياً. نقوم بالربط بين المدخلات التي تختلف عن بعضها اشتقاقياً (derivation)، إن كانت فعلاً، أو اسم فاعل، أو اسم مفعول، أو مصدرًا، أو اسم تفضيل، ... إلخ.
- دلائلياً. نستعمل الأنطولوجيا العربيّة لتفعيد وربط المعاجم مفاهيمياً، حيث نقوم بربط كل دلالة لغوية (lexical concept) لكلّ مدخلة معجمية في كلّ معجم بمفهومٍ مقابل في الأنطولوجيا (ontological concept)، وبذلك، تصبح مدخلات المعاجم مترابطة دلائلياً فيما بينها.

وبالرغم من إدراكنا للجهود الكبيرة التي يتطلّبها بناء شبكة لغوية شاملة ومحوسبة للغة العربيّة، فإنّ من شأنها - حين تتحقّق - أن تُحدث نقلة جذريّة في المعرفة اللغوية والدلالية وتؤسّس لمنهجيات جديدة للبحث النظريّ والتطبيقيّ.