

مدونات اللهجات العربية (كراسات)

<https://sina.birzeit.edu/currasat>

جُمِعَت مدونات نصية لست لهجات عامية (بما مجموعه 1.3 مليون كلمة) وأُجْرِيَ العمل على توسيم كل كلمة توسيماً صريفاً (15 سمة صرفية) وربطها بمقابلاتها في الفصحى. تشمل اللهجات العامية التي استهدفت: اللهجة الفلسطينية واللبنانية والعراقية والليبية والسودانية واليمينية، خاصة وأن هذه اللهجات هي الأقل حظاً في البحث العلمي حالياً. تهدف هذه المدونة إلى إغناء الحاسوب وتقنيات الذكاء الاصطناعي، ومساعدتها على فهم النصوص المكتوبة بالعاميات العربية، ولاحقاً تحويل العامية إلى الفصحى، إذ إن الغاية من المدونة ليست ترويجاً للعامية بتاتاً، بل لبناء تطبيقات تُمكن عامة الناس من التفاعل مع الآلة.

- التوسيمات التصريفية التي اعتمدت لتوسيم الكلمات هي: السوابق، الجذر/الساق، اللواحق، قسم الكلام، الجنس، العدد، المدخلة المعجمية العامية، المدخلة المعجمية الفصحى، تعريف إنجليزي، وغيرها.
- أكثر المستخدمين لمنصة كراسات هم الطلبة الأجانب الذين يتعلمون اللغة العربية (الفصحى والعامية).
- تشكل هذه المدونة أضخم مدونة مؤسمة توسيماً صريفاً للهجات العربية.
- جُمِعَت النصوص من مواقع التواصل الاجتماعي.
- أُجْرِيَ توسيم كل لهجة بواسطة أشخاص يتكلمون تلك اللهجة، وذلك بعد تدريبهم على منهجية رصينة.
- المدونة مفتوحة المصدر ومتاحة للبحث على منصة كراسات.
- حازت هذه المدونة على تغطيات فضائية وإذاعية وصحفية (انظر تقرير الجزيرة مثلاً).
- نُشِرَت عدة أوراق علمية وصفت المشروع من الناحية البحثية (حوسبة اللغة).

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

Mustafa Jarrar, Fadi Zaraket, Tymaa Hammouda, Daanish Masood, Martin Waehlich: **Lisan: Yemeni, Iraqi, Libyan, and Sudanese Arabic Dialect Corpora with Morphological Annotations**. arXiv, DOI 10.48550/ARXIV.2212.06468. 2022

* ورقة تصف مدونة لأربع لهجات عربية

Karim El Haff, Mustafa Jarrar, Tymaa Hammouda, Fadi Zaraket: **Curras + Baladi: Towards a Levantine Corpus**. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2022), Marseille, France. 2022

* ورقة تصف مدونة اللهجة اللبنانية وإعادة توسيم المدونة الفلسطينية

Mustafa Jarrar, Nizar Habash, Faeq Alrimawi, Diyam Akra, Nasser Zalmout: **Curras: An Annotated Corpus for the Palestinian Arabic Dialect**. Journal Language Resources and Evaluation. Pages(745-775). Volume(51), Issue(3). Springer (doi.org/10.1007/s10579-016-9370-7). 2017

الورقة الرئيسية التي تصف مدونة العامية الفلسطينية

Mustafa Jarrar, Nizar Habash, Diyam Akra, Nasser Zalmout: **Building a Corpus for Palestinian Arabic: a Preliminary Study**. Arabic Natural Language Processing Workshop, at the Conference on Empirical Methods in Natural Language Processing (EMNLP 2014). Pages(18-27). Association for Computational Linguistics. ISBN:9781937284961. Qatar. October, 2014

ورقة أولية للعمل على اللهجة الفلسطينية

في الإعلام، حول حوسبة اللهجات العربية

Corpora for Six Arabic Dialects Released

(بالعربية)



Morphologically-annotated corpora (1.3 million words) for Palestinian, Lebanese, Iraqi, Libyan, Sudanese, and Yemeni dialects are available through the [Currasat](#) portal, to enable AI to understand texts written in dialectal Arabic. Birzeit University conducted this research in cooperation with AUB and UN.



[more→](#)

Launching the Palestinian Dialect Corpus (Curras)

(بالعربية)



An online corpus for the Palestinian dialect (called Curras) was launched on January 16, 2015 by the Computer Science Department at Birzeit University. It contains about (56,000 words) of Palestinian dialect that were collected from several resources. Each word was manually annotated with 16 features.



[more→](#)

MoU with RedCrow to Cooperate in Computational Linguistics

(بالعربية)



The Computer Science Department at Birzeit University and RedCrow Intelligence signed an agreement to build smart security-altering systems. The agreement seeks to bridge the gap between academia and the industry by creating a platform for the exchange of scientific and technical expertise between both parties.



[more→](#)

تقرير قناة الجزيرة



مقابلة راديو أحيال



حوسبة العامية الفلسطينية



أمثلة (screenshots) توضح توسيمات كلمة "بلشنا"
<https://sina.birzeit.edu/currasat>



ابحث بلشنا

ع | EN

- كلمة ساق مدخلة تعريف
 فلسطينية لبنانية عراقية ليبية سودانية يمنية
 الكلمة كاملة جزء من الكلمة

حول أبحاث

5 results (1.3 secs)

تعريف	مدخلة	قسم الكلام	لواحق	ساق	سوابق	كلمة	سياق
start/begin	بَلْشْن / بدأ	فعل ماضي جمع، منكر، متكلم	نا/للماضي: فاعله متكلم جمع	بَلْشْن		بَلْشْنَا	وفلسطين بَلْشْنَا نَمْسِي
start/begin	بَلْشْن / بدأ	فعل ماضي جمع، منكر، متكلم	نا/للماضي: فاعله متكلم جمع	بَلْشْن		بَلْشْنَا	الحوثي عميل مجرم يحارب بوكالة إيران وشي معروف بس حنا بَلْشْنَا بالخون من اليمنيين اللي معنا وهم ضدنا الحوثي
start/begin	بَلْشْن / بدأ	فعل ماضي جمع، منكر، متكلم	نا/للماضي: فاعله متكلم جمع	بَلْشْن		بَلْشْنَا	بس بَلْشْنَا فيكم ما نعرف منهو الحوثي من السنني والا اليمنني الأصلي علي
start/begin	بَلْشْن / بدأ	فعل ماضي جمع، منكر، متكلم	نا/للماضي: فاعله متكلم جمع	بَلْشْن		بَلْشْنَا	من فين مصدرها نقول من الحوثي ومن آمنه قطر والا إيران. بَلْشْنَا بتحديد المرسل. زبائل الخلق تتكلم وهي تعلم أنها في برميل نفايات
start/begin	بَلْشْن / بدأ	فعل ماضي جمع، منكر، متكلم	نا/للماضي: فاعله متكلم جمع	بَلْشْن		بَلْشْنَا	كل شوي مرسل صاروخ علينا ولا أحد تكلم!!هذي والله العقوبة صدق بَلْشْنَا بهالسلق الي مايناجون إلا علينا بيللا_حديد