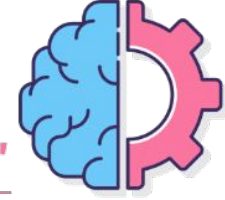




# اسأل نور Ask Nour



"نور" روبوت دردشة وحُزمة برمجية معجمية للغة العربية!

مشروع "اسأل نور": روبوت دردشة وحُزمة برمجية معجمية للغة العربية!  
<https://www.asknour.com>

مقدم إلى : برمجان العربية أحد مشاريع مجمع الملك سلمان العالمي للغة العربية !

مقدم من قبل: علي صالح وجمال الحلواني





سيركز المشروع على عدة مجالات في تحدى المعجم وهي:

- استخراج الأصل المعجمي Lemma (الكلمة الأصلية التي تمثل المدخل المعجمي).
- استخلاص العلاقات الدلالية آلياً (الترادف حالياً)
- التعرف على أقسام الكلام (POS)
- التحليل الصرفي
- بناء مجاميع وقواعد معرفية للتصاحب اللغوي والتعابير الاصطلاحية.
- استخراج الميزان الصرفي.
- استخراج تشكيلات ومشتقات متنوعة للكلمات.



## أهداف المشروع:

- توفير مكتبة برمجية بلغة بايثون تركز على معالجة الكلمات في اللغة العربية وبالذات في المجالات المذكورة سابقاً وهذه مخصصة للتقنيين و الباحثين والمطورين.
- توفير واجهة ومريحة للتعامل مع المكتبة البرمجية وكان برأينا أفضل ما يمكن تقديمه هو روبوت دردشة مبني فوق هذه المكتبة ويُجيب على أسئلة المستخدمين.
- تشجيع المبرمجين والباحثين وغيرهم (من خلال المكتبة) على بناء تطبيقات تدمج اللغة العربية وذلك من أجل زيادة المحتوى العربي والعمل به.
- توفير نماذج برمجية كاملة وسهلة الاستخدام والدمج على شكل python notebooks لكافة الخدمات التي تقدمها "نور".



هنالك نقص في الأدوات والتقنيات الداعمة للغة العربية وهناك صعوبة في دمجها بالتطبيقات.

المشكلة أن ما هو متوفر يحتاج منك الكثير من العمل ليتم دمجها في برامجك وتطبيقاتك وبعضها يحتاج لشروط كثيرة ليعمل وبعضها لا يعمل أصلا لأنه قديم.

طبعا أنا هنا لا أقلل أبدا من جهود غيري بالعكس هناك الكثير من الزملاء الذين يخدمون العربية ولهم كل تقدير.

المشكلة الأساسية أننا ننسى موضوع الدمج وسهولة الاستخدام وهذا يؤثر سلبا على تبني اللغة العربية وتبني التقنيات الداعمة لها في المشاريع والمبادرات

ومن هنا جاءت الفكرة في توفير مكتبة برمجية سهلة الاستخدام وبناء واجهة "روبوت دردشة" باستعمالها لإثبات الفاعلية والجدوى.



الحل المقترح "اسأل نور" هو مكتبة برمجية عصرية يتم إستعمالها من خلال لغة البايثون لخدمة المعاجم واللغة العربية واثبات فاعليتها بتوفير روبوت دردشة مبني باستعمالها.

**ملاحظة:** الحل المقترح هو مبني على نواة معرفية من مشاريعي الشخصية السابقة حيث سأقوم باستعمال القواعد المعرفية المتوفرة لدي لبناء نظام جديد متخصص فقط في المعاجم بلغة بايثون وعلى أن يتمتع النظام الجديد بما يلي:

- توفير نماذج برمجية كاملة وسهلة الاستخدام والدمج على شكل python notebooks
- توفير واجهة تطبيق ويب (ملائم للهواتف أيضاً) على شكل روبوت دردشة اسمه "اسأل نور"
- توفير الحزم البرمجية المصاحبة على github ليتم سحبها بسلاسة ودمجها في أي مشروع بحثي او برمجي.



بعض تقنيات الذكاء الاصطناعي التي سوف يتم استخدامها في تطوير بعض الأجزاء:

- بالنسبة للتعرف على أقسام الكلام سيتم استخدام موارد معرفية من \*Universal Dependencies وبالاخص للغة العربية لتدريب خوارزمية على فهم اقسام الكلام باللغة العربية (مثال تدريب خوارزمية مثل Limited-memory BFGS للتعرف على أقسام الكلام).
- استعمال موديل تعلم عميق مثل ( Keras LSTM ) بالإضافة الى دعم التجارب بقوائم خاصة من wikipedia للتحليل أقسام الكلام.



- سنقوم باستعمال موديل التعلم العميق مثل \*CAMELBERT-Mix NER Model وهو مودل مبني على BERT بالأساس لإستخراج التعابير الاصطلاحية وتدريب الموديل لتحسين دقته ودمجه في الأدوات.
- 
- بالإضافة من الممكن استعمال موديل اخر مثل \*\*Arabic NER Model using Flair Embeddings وعمل دمج بينهم : ensembling
- بالنسبة لإستخراج الأصل المعجمي كنت قد دربت خوارزمية في السابق ولدي مصدر معرفي مبني والذي من الممكن تحسينه بتدريب آخر باستعمال Universal Dependencies او اي مصدر معرفي آخر.

\* <https://huggingface.co/CAMEL-Lab/bert-base-arabic-camelbert-mix-ner>

\*\* <https://huggingface.co/megantosh/flair-arabic-multi-ner>





يتوفر لدى عدد من القواعد المعرفية والتي استطيع منها بناء وتحسين ما يتم تطويره باستمرار واذكر منها :

- العديد من البيانات تم تجميعها مسبقاً خلال عملي في مجال اللغة العربية\* وبعضها من خلال أبحاثي السابقة المنشورة في مجلات علمية و مؤتمرات بحثية عربية ودولية واستطيع توفير هذه الأبحاث عن طلبها.
- العديد ايضاً من البيانات تم بنائها مسبقاً في مشاريعي الشخصية والتي طورتها بمجهودي الفردي اعتمدت وقتها على بعض الكتب والتجميع اليدوي واستطيع توفير هذه البيانات عند طلبها.
- قوائم الكلمات تم جمعها من خلال مسح اللويكيبيديا و مسح لبعض الجرائد العربية مثل جريدة القدس الفلسطينية في وقت سابق.



- قوائم التعابير الاصطلاحية تم جمعها من خلال خوارزمية طورتها لربط الكلمات ببعضها البعض بالاعتماد على التصنيفات في الويكيبيديا العربية.
- بعض المترادفات الاساسية تم جمعها يدويا من عدة مصادر منها قاموس الطالب في المرادفات والأضداد\*.
- الكثير من الجذور والكلمات ذات العلاقة بها تم جمعها من الويكيبيديا العربية حيث تتوفر قوائم هناك وتم بناء جزء من القواعد المعرفية بها\*\*.

\* <https://books-library.net/free-928746457-download>

\*\* [https://en.wiktionary.org/wiki/Category:Arabic\\_lemmas](https://en.wiktionary.org/wiki/Category:Arabic_lemmas)



- قوائم الخاصة في التعرف على أقسام الكلام سيتم اعتماد مصادر مثل Universal Dependencies\* والتي تحتوي على بيانات للتدريب للموديلات العميقة والإحصائية.
- ايضا سيتم استعمال \*\*Hugging Face Datasets لجلب قواعد بيانات معرفية جديدة داعمة عن الحاجة.

\* <https://universaldependencies.org/>

\*\* <https://huggingface.co/docs/datasets/index>



### التركيز هنا سيكون باتجاهين:

الاتجاه الأول: توفير حزم برمجية على شكل أكواد برمجية بلغة البايثون ضمن المميزات التالية :

- حزمة برمجية لاستخراج الأصل المعجمي
- حزمة برمجية لاستخراج المترادفات
- حزمة برمجية للتحليل الصرفي.
- حزمة برمجية لبناء قوائم التعابير الاصطلاحية.
- حزمة برمجية لإستخراج الميزان الصرفي للفعل وتشكيلاته.



**الإتجاه الثاني:** عن طريق بناء روبوت دردشة اطلقنا عليه اسم "نور" (وهو اسم زوجتي) وهذا الروبوت سيكون قادراً على التفاعل مع الجمهور بما يخص الحُزم البرمجية السابقة وتجربتها.

ستتوفر هذه الحُزم مكتبة برمجية مع امثلة شاملة للإستعمال والدمج في مشاريع مختلفة وتتوفر على شكل Python Notebooks للجميع وهذا سوف يساعد على سهولة دمجها في أبحاث وبرامج الآخرين مما يحفز ويساعد في دفع عجلة التطوير في حوسبة اللغة العربية.



قائد الفريق: علي صالح

التعليم: ماجستير إدارة أعمال وبكالوريوس هندسة أنظمة حاسوب.

التخصص الدقيق: متخصص في مجال استرجاع المعلومات ومعالجتها و متخصص في التحليل والمعالجة الرقمية للغات الطبيعية وخاصة اللغة العربية - ١٤ سنة خبرة.

لدي العديد من المنشورات (أكثر من ١٤ بحث محكم في المجلات العلمية والمؤتمرات) وحاصل على العديد من الجوائز في المجال.

مصنف كخبير مسابقات ونقاش معتمد على موقع Kaggle

للمزيد من المعلومات:

<https://asalhi.info/>

<https://www.linkedin.com/in/asalhi85>



عضو فريق: جمال الحلواني

التعليم: بكالوريوس لغة عربية .

التخصص الدقيق: تعليم وتدرّيس لغة عربية و دراما في سياق التعليم.

سنوات الخبرة : ٦ سنوات



هذه بعض الصور التوضيحية - ليس بالضرورة أن يكون المنتج النهائي مطابق تماماً

Ask Nour اسأل نور  
"نور" روبوت دردشة ! وهي تعرف الكثير عن اللغة العربية !



ما هو وزن كلمة يلعب ؟

[يُفَعِّلُ، يُفَعِّلُ، يُفَعِّلُ، يُفَعِّلُ، يُفَعِّلُ، يُفَعِّلُ، يُفَعِّلُ، يُفَعِّلُ، يُفَعِّلُ، يُفَعِّلُ، يُفَعِّلُ، يُفَعِّلُ، يُفَعِّلُ، يُفَعِّلُ، يُفَعِّلُ]



ما هي مرادفات كلمة بيت يا نور ؟؟

[إقامة، مأوى، مسكن، دار سكن، بيت، أهل، مأوى، مسكن، مَبَاة، مُدَبَّر، رَنْع، منزل، موطن، شقة، مُبَيَّت، شقة،

بَيْت، مَرْسُوم، مقر بيت]



اسأل نور من هنا ...







هذه بعض الصور التوضيحية - ليس بالضرورة أن يكون المنتج النهائي مطابق تماماً



وماذا عن مرادفات كبير ؟

[ضخم, مهم, جَبَّار, كَثِير, عَظِيم, جَرِيم, كَبِير, جسيم, عظيم, كثير, صَخْم, هائل, جَسِيم, هائل, بالغ, كبير, كبرى, جَرَّار, هام, شديد, غَزِير, عَمَلَق]



ووزن كلمة يتنافسون ؟

[يَتَفَاعَلُونَ, يُتَفَاعَلُونَ, يتفاعلون]



اسأل نور من هنا ...





مثال لجزء من التحليل الصرفي (من المكتبة البرمجية المعجمية - المدخل : يقرؤون)

```
input_word = "يقرؤون"  
args = ['../input/asknour/asknour_verbs_tasreef.jar', input_word]  
  
result = jarWrapper(*args)  
data = json.loads(json.loads(json.dumps(jarWrapper(*args)))[0])  
#print(data)  
print ('المضارع - هو : ' + data['present_tense']['howa'])  
print ('الماضي - نحن : ' + data['past_tense']['nahno'])  
print ('فعل الأمر - انت : ' + data['imperative_tense']['anti'])  
print ('اسم المفعول : ' + data['participles_passive']['value'])  
print ('الفعل المضارع المجزوم - هن : ' + data['present_tense_jussive']['hunna'])
```

المضارع - هو : يَقْرَأُ  
الماضي - نحن : قَرَأْنَا  
فعل الأمر - انت : اقْرَأْ  
اسم المفعول : مَقْرُوء  
الفعل المضارع المجزوم - هن : يَقْرَأْنَ



مثال لنتيجة متوقعة من استدعاء المكتبة البرمجية المعجمية لتصريف الفعل "قرأ" - طبعاً هنالك حرية هنا في تصنيف واستخلاص ما يرغب المطور من البيانات المسترجعة

```
"verb": "قَرَأَ",
"passive_past_tense": { + },
"passive_present_tense_confirmed_heavy": { -
  "howa": "يُقْرَأُنْ",
  "hum": "يُقْرَوْنَ",
  "hunna": "يُقْرَأُنَّ",
  "huma_female": "تُقْرَأُ",
  "antoma_female": "تُقْرَأُ",
  "antoma_male": "تُقْرَأُ",
  "nahno": "تُقْرَأُ",
  "anti": "تُقْرَبُ",
  "antom": "تُقْرَوْنَ",
  "antona": "تُقْرَأُنَّ",
  "ana": "أُقْرَأُ",
  "heya": "تُقْرَأُ",
  "anta": "تُقْرَأُ",
  "arabic_name": "الفعل المضارع المؤكد الثقيل المبني للمجهول",
  "huma_male": "يُقْرَأُ"
},
"src": { -
  "arabic_name": "المصدر",
  "value": "قِرَاءَة"
},
},
```

```
{ -
  "present_tense": { + },
  "passive_present_tense_jussive": { -
    "howa": "يُقْرَأُ",
    "hum": "يُقْرَوُا",
    "hunna": "يُقْرَأُنَّ",
    "huma_female": "تُقْرَأُ",
    "antoma_female": "تُقْرَأُ",
    "antoma_male": "تُقْرَأُ",
    "nahno": "تُقْرَأُ",
    "anti": "تُقْرَبُ",
    "antom": "تُقْرَوُا",
    "antona": "تُقْرَأُنَّ",
    "ana": "أُقْرَأُ",
    "heya": "تُقْرَأُ",
    "anta": "تُقْرَأُ",
    "arabic_name": "الفعل المضارع المجزوم المبني للمجهول",
    "huma_male": "يُقْرَأُ"
  },
  "src": { -
    "arabic_name": "المصدر",
    "value": "قِرَاءَة"
  },
},
```

شكرا لكم - مع فائق الإحترام والتقدير



Ask Nour اسأل نور



"نور" روبوت دردشة وخُزْمة برمجية معجمية للغة العربية!

<https://www.asknour.com>