

لهم إني
أعوذ بِكَ مِنْ أَنْ يَأْتِيَ
نَفْسٌ مِّنْ
مَا نَهَى

مسار المعجم

- إستخراج الأصل المعجمي

- التعرف على أقسام الكلم

المشكلة المستهدفة

في اللغات السامية مثل اللغة العربية مهمة استخراج الأصل المعجمي ليست مهمة شكلة بسبب خصائصها التصريفية العالية.

حيث تكمن حاجة الأصل المعجمي لاستخدامه في مراحل أخرى لمعالجة اللغات الطبيعية، لتجنب التفريق بين عدة كلمات غالباً ما تكون مشتقة من بعضها البعض

مجالات الحاجة إلى أصل الكلمة المعجمي:

٥ استرجاع المقامات [IR]

٤ استخراج العبارة الرئيسية

٣ التلخيص

٢ الترجمة الآلية

١ الفهرسة



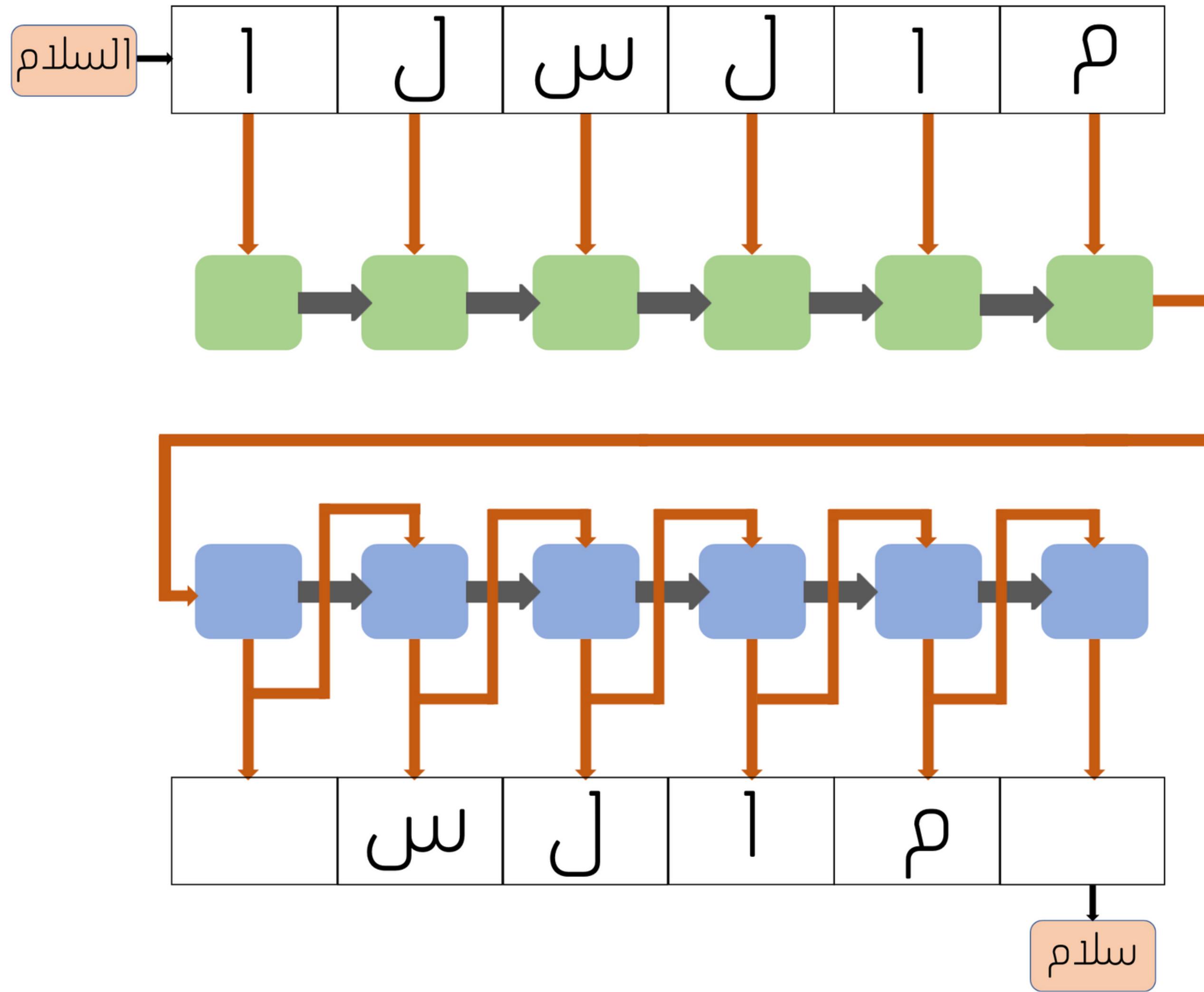
الحل المستخدم

Seq2seq ML

ما هي طريقة تنفيذ الحل؟

تم استخدام الترجمة الآلية العربية NMT

عادةً عند استخدام الترجمة الآلية العميقه يتم تحويل كل كلمة إلى رمز مقابل ليسهل فهمها على الآلة، لكن في حالة مشروع أصل تم تحويل كل حرف إلى رمز، ليتمكن النموذج من استخراج جميع أصول الكلمات حتى التي لم يسبق لها رؤيتها.



البيانات المستخدمة

مجموعة بيانات عبارة عن الكلمة إضافة إلى أصلها المعجمي
إضافة إلى تحديد جزئها من الكلام

```
{"Word": "ذهب", "Lemma": "ذهب", "POS": "verb"},
```

مصادر البيانات:

مفردات (PADT 1.0) وتحوي على عدة معلومات تمأخذ ال lemma وال POS منها

<https://lindat.mff.cuni.cz/repository/xmlui/handle/11858/00-097C-0000-0001-4872-3#>

مفردات القرآن الكريم كاملاً مع أصلها وال POS الخاص بها

<https://corpus.quran.com/lemmas.jsp>

أصحاب فريق أصل

عامم الشاع - (تعلم آلة+لغوي)

عبدالله الغزي - لغوي

عبدالرحمن الركف - مبرمج

نموذج أولي تجاري للمشروع

يرجى إدخال النص أو الكلمة ليتم استخراج أصل الكلمة المعجمي

يُشار إلى أن اللغة العربية يتحدثها أكثر من مليون نسمة ويتوسط متحدثوها في المنطقة المعروفة باسم الوطن العربي بالإضافة إلى العديد من المناطق الأخرى المجاورة مثل الأهواز وتركيا وتشاد والسنغال وغيرها هي اللغة الرابعة من لغات منظمة الأمم المتحدة الرسمية ست

النص بعد المعالجة:

أشار إلى أن لغة عربي تحدث أكثر من مليون نسمة توزع متحدث في منطقة معروف اسم وطن عربي إضافة إلى عديد من منطقة أخرى مجاور مثل أهواز تركيا تشاد سنغال وغيرها هي لغة رابع من لغة منظمة أمم متحد رسمي ست

استخدام ذات النص تجريبي باستخدام FARASA وهي أحد أكبر معالجات اللغة العربية

يُشار إلى أن اللغة العربية يتحدثها أكثر من 422 مليون نسمة ويتوزع متحدثوها في المنطقة المعروفة باسم الوطن العربي بالإضافة إلى العديد من المناطق الأخرى المجاورة مثل الأهواز وتركيا وتشاد والسنغال وإريتريا وغيرها. وهي اللغة الرابعة من لغات منظمة الأمم المتحدة الرسمية السنت.

Lemmatize Text

أشار إلى أن لغة عربي تحدث أكثر من 422 مليون نسمة توزع متحدثوها في منطقة معروف اسم وطن عربي إضافة إلى عديد من منطقة آخر مجاور مثل أهواز تركيا تشاد سنغال إريتريا غير. هي لغة رابع من لغة منظمة أمم متحدة رسمي سنت .

نلاحظ الخطأ في استخراج أصل الكلمة (متحدثوها)
بينما قام النموذج السابق باستخراجها بشكل صحيح.

الله
لهم
أنت ملائكة
أنت ربنا
أنت رب العالمين