

Lemmatizer مبني على المعجم العربي

مسار: تحدي المعجم المجال: استخراج الاصل المعجمي



الهدف من المشروع

1. إثراء المحتوى العربي في مجال تعلم الآلة ومعالجة اللغات الطبيعية من خلال انشاء وبرمجة مكتبة في لغة بايثون تحتوي على **lemmatizer** لإعادة الكلمات الى شكلها الأساسي الأولي مبني على المعجم العربي بشكل الي بالكامل.

2. دعم الباحثين في مجال تعلم الآلة ومعالجة اللغات الطبيعية **NLP** باللغة العربية ورفع دقة نماذج تعلم الآلة في هذا المجال.



وصف المشكلة



وجدنا ان معظم الباحثين في مجال معالجة اللغات الطبيعية باللغة العربية يعانون من مشكلة ضعف المكتبات العربية في هذا المجال، مما يسبب ضعف دقة نماذج تعلم الالة.

الحل المقترح

بناء مكتبة في لغة بايثون مبنية على المعجم العربي، تستطيع من خلال هذه المكتبة استخدام الـ **Lemmatizer** والذي بدوره سيعيد الكلمات في النصوص الى شكلها الحقيقي الأولي.

بناء المكتبة سوف يكون بشكل الي من خلال ربط الكلمة مع اصلها الأولي في المعجم مثال: (الكلمة : يلعب، الأصل : لِعِبَ)، ومن ثم يتم استخدام هذا الـ **Lemmatizer** لارجاع الكلمات في النصوص لأصلها الأولي ليساعد في رفع كفاءة نماذج تعلم الآلة باللغة العربية.



البيانات



سيتم استخدام المعجم الإلكتروني وبرمجة المكتبة بناء عليه، وسيتم اختبار كفاءته على بيانات
Arabic Sentiment Analysis 2021 @ KAUST والمتوفرة على **Kaggle** بشكل مفتوح
المصدر باستخدام أكثر من نموذج تعلم الآلة مثل:
SVM, MLP, Decision Tree, Random forest.

القيمة المضافة



اثراء المحتوى العربي ومساعدة الباحثين من خلال
انشاء مكتبة تساعد في رفع كفاءة تنظيف البيانات
ومعالجتها قبل ادخالها الى نموذج تعلم الالة والتي
بدورها سوف ترفع دقة النموذج المبني على
النصوص العربية.



فريق العمل

عبدالرحمن منصور المالكي
التخصص
(الذكاء الاصطناعي)

عبدالرحمن يونس المالكي
التخصص
(الأمن السيبراني)

فواز مبارك الشيعي
التخصص
(الذكاء الاصطناعي)