

كيف نُعلِّم الذكاء الاصطناعي

فهم اللهجة الشامية الدمشقية؟



يهدف هذا المشروع إلى توثيق
المصطلحات الشامية ضمن إطار
لغوي قابل للاستخدام في
تطبيقات الذكاء الاصطناعي
ومعالجة اللغة العربية.





لماذا هذا المشروع مهم؟

1 بعد السنوات الأخيرة، ازداد اهتمام العالم باللهجة والثقافة السورية.

2 الحاجة إلى بيانات لغوية دقيقة اللهجات العربية لاستخدامها في البحث، التعليم، والذكاء الاصطناعي.

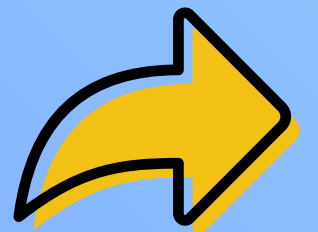
3 معظم أدوات NLP لا تتعامل مع اللهجات جيداً بسبب نقص البيانات المنظمة.





الهدف النهائي

1. بناء قاعدة بيانات لغوية منظمة تشمل:
2. المصطلح الشامي
3. كتابته الموحدة (Normalized)
4. معناه العربي الفصيح
5. دوره النحوي (POS)
6. السياق الذي يظهر فيه
7. تخزينه في قاعدة PostgreSQL للاستخدام المستقبلي





مخطط العمل (Pipeline)

1

Whisper ASR

2

Human Review

3

Normalization

4

Tokenization

5

Normalization

6

تخزين البيانات في
PostgreSQL

05



من الصوت إلى نص

استخدام Whisper لتحويل التسجيلات إلى نص



مراجعة بشرية لتصحيح الأخطاء



توحيد الكتابة (Normalization)





معالجة النص

1

Tokenization باستخدام CAMEL Tools

2

إزالة التكرار الحرفي والضبابي

3

إضافة المعاني

4

تحليل نحوي باستخدام Stanza



07

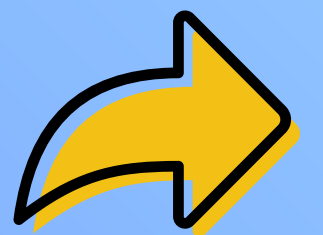




النتائج النهائي

قاعدة بيانات لغوية منظمة للهِجة الشامية
تدعم:

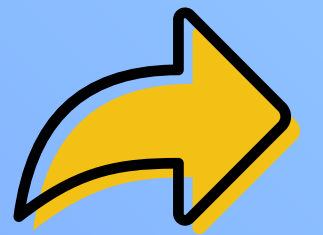
1. تدريب نماذج NLP
2. بناء Chatbots
3. تطوير أدوات تعليمية
4. دراسات لغوية للهجات السورية



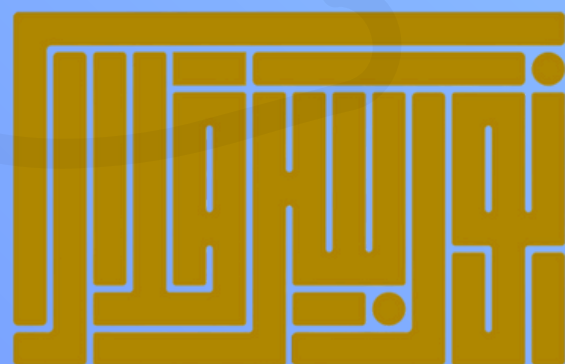
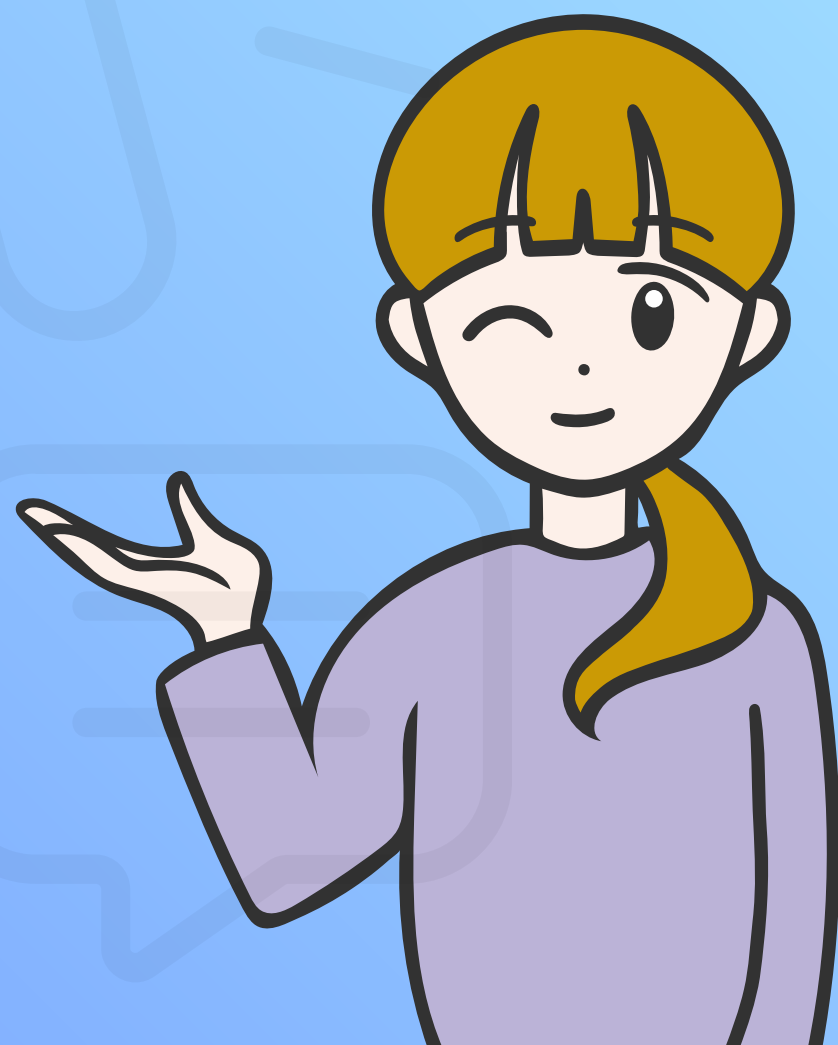


برأيك

ما هي المصطلحات الشامية التي تعتقد
أنها منتشرة بكثرة، ولكن لم يتم تضمينها
بعد في أدوات الذكاء الاصطناعي؟



تابعني لنتساعد على نشر المعلومات التقنية باللغة العربية



Eng. Nour Berkdar