

Farasa Arabic NLP Add-in to Microsoft Office

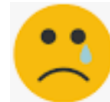


Mentors: Hamdy Mubarak, Kareem Darwish, Ahmed Abdelali {hmubarak, kdarwish, aabdelali}@hbku.edu.qa

Motivation: There are **millions** of Arab users of Microsoft Office (Word, Excel, etc.). **Integrate Farasa** allows users to benefit from its capabilities (diacritization for example which doesn't exist in MS-Word) in a **convenient** way without losing document **format and layout** or having size limits.

Background: Office Add-ins platform allows us to build solutions that **extend Office applications** and interact with content in Office documents. We will interact with Word, Excel, PowerPoint, OneNote, Project, and Outlook.

- Solutions can run in Office across **multiple platforms** (Office for Windows, Office Online, Mac and iPad)
- **Farasa Arabic NLP** tools perform many tasks like: stemming, diacritization, spell checking, etc. and can be called using a command line as a Java package or from the online interface.
- Farasa assumes **plain text** and doesn't keep format (line boundaries, fonts, colors, tables, etc.)



Diacritization التشكيل www.farasa.qcri.org

التَّدرِيبَاتُ البَدَنِيَّةُ تَقِي حَظَرَ الزَّهَائِمِرِ تُسَاعِدُ التَّدرِيبَاتُ البَدَنِيَّةُ اليَوْمِيَّةُ عَلَى حَفْضِ حَظَرِ .
الإصَابَةُ بِمَرَضِ الزَّهَائِمِرِ بِشَكْلِ وَاضِحٍ . التَّجْرِبَةُ مَلاحَظَاتُ . دِرَاسَةُ النِّشَاطِ الجَسَدِيِّ لِ 716
مُسِنًا تَبَيَّنَ لِلبَاحِثِينَ أَنَّ الأَشْخَاصَ الَّذِينَ كَانُوا الأَقَلَّ حَرَكََةً مِنْ بَيْنِ هَؤُلَاءِ أُصِيبُوا بِالزَّهَائِمِرِ
بِنِسْبَةٍ 2.3 مَرَّةً أَعْلَى مِنْ نِسْبَةِ 10% الأَكْثَرُ نَشَاطًا

• The demo is confined to process only three sentences per request. each sentence shouldn't exceed 20 words.
• The length of the text to be processed should be within 400 characters.



Farasa

Diacritize التشكيل النص
Correct تصحيح الأخطاء
Index فهرسة النص
Keywords الكلمات المفتاحية

التَّدرِيبَاتُ البَدَنِيَّةُ تَقِي حَظَرَ الزَّهَائِمِرِ
تُسَاعِدُ التَّدرِيبَاتُ البَدَنِيَّةُ اليَوْمِيَّةُ عَلَى حَفْضِ حَظَرِ
الإصَابَةُ بِمَرَضِ الزَّهَائِمِرِ بِشَكْلِ وَاضِحٍ

ملاحظات	التجربة
تَبَيَّنَ لِلبَاحِثِينَ أَنَّ الأَشْخَاصَ الَّذِينَ كَانُوا الأَقَلَّ حَرَكََةً مِنْ بَيْنِ هَؤُلَاءِ أُصِيبُوا بِالزَّهَائِمِرِ بِنِسْبَةٍ 2.3 مَرَّةً أَعْلَى مِنْ نِسْبَةِ 10% الأَكْثَرُ نَشَاطًا	دِرَاسَةُ النِّشَاطِ الجَسَدِيِّ لِ 716 مَسْنًا
..	..



❖ Solution

- **Integrate** Farasa diacritizer to MS-Office
- **Test** with different layouts and platforms.
- Allow user to click on a word to get its **linguistic info**: root, pronunciation...
- Integrate other components, ex: **spell checker** and test complex operations like: Replace All, Add to Dictionary, etc.

❖ Skills

- **Good programming skills** (JavaScript, Java)
- **Familiarity with HTML, CSS**
- **Familiarity with GitHub** (source control)