

Project Title

Farasa Arabic NLP Add-in to Microsoft Office

Project Description:

There are millions of Arab users of Microsoft Office (Word, Excel, etc.). Integrate Arabic NLP tools (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.

Please have a look at the presentation slide below!

Duties/Activities:

Integrate a Java package tool to MS-Office (MS-Word as a start), test for different types of formats, layouts and platforms, and package the add-in such that it can be installed easily.

Required Skills:

Good programming skills (JavaScript, Java) and familiarity with HTML, CSS

Preferred Intern Academic Level:

Under-grad students are welcomed!

Learning Opportunities:

Learn about [Office Add-ins platform](#) which 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). Also, learn about Arabic processing tools and packaging (ex: installation)

Expected Team Size: *it is preferable to have team projects*

2 to 3

Mentors

Name:	email:
Hamdy Mubarak	hmubarak@hbku.edu.qa
Dr. Kareem Darwish	kdarwish@hbku.edu.qa
Dr. Ahmed Abdelali	aabdelali@hbku.edu.qa

Farasa Arabic NLP Add-in to Microsoft Office

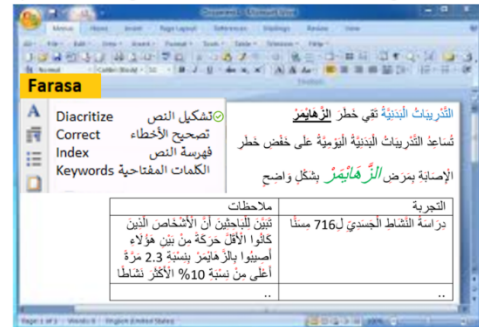
- **Mentors:** Hamdy Mubarak, Kareem Darwish, Ahmed Abdelali
- **Motivation:** There are millions of Arab users of Microsoft Office (Word, Excel, etc.). Integrate Farasa allows users to benefit from its capabilities (diacritization for example which doesn't exist in MS-Word) in a **convenient** way without 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.)



➤ Solution

- **Integrate** Farasa diacritizer to MS-Office
- **Test** with different layouts and platforms
- Integrate other components, ex: **spell checker** and test complex operations like: Replace All, Add to Dictionary, etc.



➤ Skills Needed:

- Good programming skills (JavaScript, Java)
- Familiarity with HTML, CSS

