



Mustafa Jarrar with some of the 150 Arabic dictionaries that were scanned into the Arabic Ontology online database (Photo: Eyad Jadallah/ Birzeit University).

The mother of all Arabic dictionaries makes its debut online

📅 POSTED: SATURDAY 12.01.2018 12:00 PM 🔄 REVISED: FRIDAY 11.30.2018 12:00 PM
 👤 THEAANEWS (HTTPS://WWW.ARABAMERICANNEWS.COM/AUTHOR/ADMINUSERX/)
 🗂️ COMMUNITY (HTTPS://WWW.ARABAMERICANNEWS.COM/CATEGORY/COMMUNITY/), WORLD (HTTPS://WWW.ARABAMERICANNEWS.COM/CATEGORY/WORLD/)

WEST BANK — A computer science professor at Birzeit University has built an online tool called Arabic Ontology that is both a comprehensive dictionary of Arabic and a system that will enable the creation of new Arabic-language software, such as better machine translation.

Mustafa Jarrar ([//www.jarrar.info/](https://www.jarrar.info/)), an associate professor in the department of computer science at Birzeit, in the West Bank, has worked for eight years to create the new tool ([//www.birzeit.edu/en/news/birzeit-announces-linguistic-search-engine-and-arabic-ontology/](https://www.birzeit.edu/en/news/birzeit-announces-linguistic-search-engine-and-arabic-ontology/)). It functions not only as a searchable lexicon of Arabic that can be used as both a dictionary and a thesaurus, but also as a logical system — an ontology — that recognizes the unique characteristics of the Arabic language. It can find relationships between the meanings of Arabic words in a natively Arabic way for the first time.

The Arabic Ontology offers the prospect of more precise results from Google searches in Arabic, better online translation of Arabic text and new insights into the language for students and scholars of Arabic literature.

The new tool, which is available for personal use for free at [//ontology.birzeit.edu/](http://ontology.birzeit.edu/) ([//ontology.birzeit.edu/](http://ontology.birzeit.edu/)), was made public in a ceremony at Birzeit on September 25. The copyright is owned by Birzeit University.

“This is the first search engine of its type for a single language,” Jarrar said.

That is, the search engine offers results from 150 Arabic dictionaries.

“Imagine you had the Oxford English Dictionary, the Merriam-Webster dictionary and all the others collected in one place, all integrated and unified in one database,” he said.

To create the combined Arabic lexicon, the contents of 150 Arabic dictionaries had to be manually entered into a database. This was painstaking work. At first, Jarrar tried capture information from books by using a scanner. But to extract the useful data from the books required optical character recognition software, known as OCR, that could read Arabic.

“I tried to use OCR, but it didn’t work,” Jarrar said. Arabic OCR software is still so poor, he said, that “the amount of corrections you have to do is more work than if you entered the text from scratch, manually.”

Instead, Jarrar said, he crowdsourced the task to Birzeit University students.

Birzeit requires students to perform 120 hours of community service before they can graduate. The program’s goal is to better connect the university with Palestinian society outside the campus. Typically, Birzeit students doing community service will pick olives on local farms, or help old people in their homes.

The Birzeit administration considered that working on Jarrar’s project was a suitable activity for community service, because of its value to Arab culture and society as a whole, Jarrar said, so students could fulfill their community service duty by typing the contents of pages of Arabic dictionaries into his database. To improve accuracy, he gave the same page to more than one student to transcribe.

Eventually, he selected students who could do the work to a high standard, without mistakes. He engaged these students as paid workers. The work took eight years to complete.