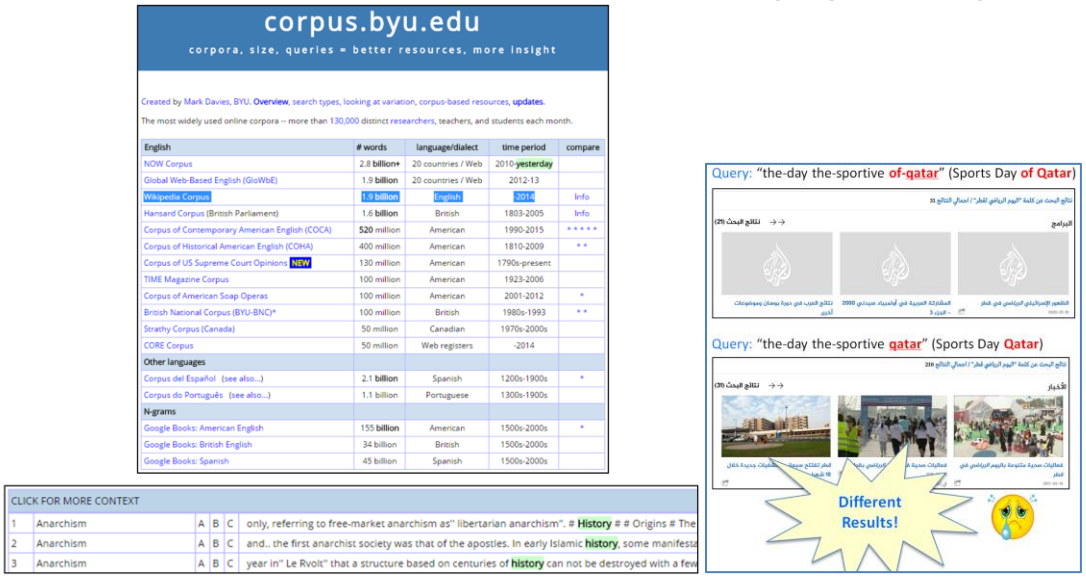
**Project Title: Enhance Arabic Search**

**Project Description:**

**Search Engines** (SE's) should consider Arabic rich morphology. SE's in even famous Arabic sites (like Aljazeera.net) don't fully support Arabic characteristics! (Search by stem or lemma, root, derivatives, etc.)

Solution: As an example for a morphological SE, we will:

- ✓ Build a site to search in **Arabic Wikipedia** (similar to corpus.byu.edu, from Brigham Young University, which has texts in English, Spanish, and Portuguese).
- ✓ Call **QCRI's Farasa** tools to extract for each word: its root, stem, etc.
- ✓ Use **Solr** search engine to index words and their morphological info.
- ✓ Build **interface**!
- ✓ Search for **complex queries** (ex: **What are the adjectives that follow a certain noun?**) which are important for linguistic study and language learning.




**Required Skills:**

- Good programming skills (Python/Java/C#/C++).
- Understanding Arabic is a plus but not a must!

**Expected Team Size:**

One to two students

**Mentors:** Hamdy Mubarak, Kareem Darwish, and Ahmed Abdelali


| **Name:** | **e-mail:** |
|---|---|
| Hamdy Mubarak | hmubarak@hbku.edu.qa |
| Kareem Darwish | kdarwish@hbku.du.qa |
| Ahmed Abdelali | aabdelali@hbku.edu.qa |