

Arabic Graph Knowledge Base

Project Description:

Knowledge bases can be represented as RDF (Resource Definition Framework) Graphs. Each RDF triple represents a relationship between a subject entity and an object. Hence, RDF graphs provide a structured way to store and process any kind of information. RDF knowledge bases is the source of data you see in the information boxes displayed when you search google for an entity or when a Wikipedia page is displayed.

There are numerous public RDF datasets on the cloud (DBpedia, Freebase, Wikidata). However, they all lack enough Arabic content. In QCRI, we have huge Arabic content that can be leveraged to enrich the RDF structured Arabic content on the cloud.

Duties/Activities:

- Explore current RDF knowledge base graphs and extract meaningful stats and insights
- Translating unstructured Arabic content into RDF triples to build new Arabic datasets

Required Skills:

- Awareness of Graph representation and traversal fundamentals
- Mastering at least one high level programming languages (Python, Java ... etc)
- Machine Learning background

Preferred Intern Academic Level:

- 1) Senior Students
- 2) Junior Students

Learning Opportunities:

- RDF
- SPARQL (graph query language)
- Graph Databases

Expected Team Size: 1+

Mentors

Name: Abdurrahman Ghanem

email: abghanem@hbku.edu.qa