

## **Project Title: *Make the Google of the News World***

**Project Description:** Have you ever wondered how Google and other search engines work under the hood? Our Tanbih project, <https://www.tanbih.org/>, collects news articles from sources from all around the world and provides intelligent tools to analyse them, such as bias and propaganda analysis, stance detection of media outlets with respect to specific topics and many more.

We want to extend Tanbih by allowing users to make complex searches on news articles using also our tools as filters. We need to explore open source search engines projects, such as Solr, Elastic Search and Lucene to index the data efficiently and make it available to the users of the system.

More about the project: <http://tanbih.qcri.org/>

You can try the present system: <https://www.tanbih.org/>

**Duties/Activities:** (depending on skills and interests)

- analysis of open source search engines
- data management
- user interface design

**Required Skills:**

- basic programming
- data analytics

**Preferred Intern Academic Level:**

- All levels: PhD, MSc, undergrad

**Learning Opportunities:**

- You will learn how to build a search engine and how to efficiently organise large amounts of data, how to interface with a SQL database.

**Expected Team Size:** *1-2 people*

### **Mentors**

Name:	Mr. Yifan Zhang	email: <a href="mailto:yzhang@hbku.edu.qa">yzhang@hbku.edu.qa</a>
	Dr Michael J. Aupetit	email: <a href="mailto:maupetit@hbku.edu.qa">maupetit@hbku.edu.qa</a>
	Dr Giovanni Da San Martino	email: <a href="mailto:gmartino@hbku.edu.qa">gmartino@hbku.edu.qa</a>
	Dr Preslav Nakov	email: <a href="mailto:pnakov@hbku.edu.qa">pnakov@hbku.edu.qa</a>