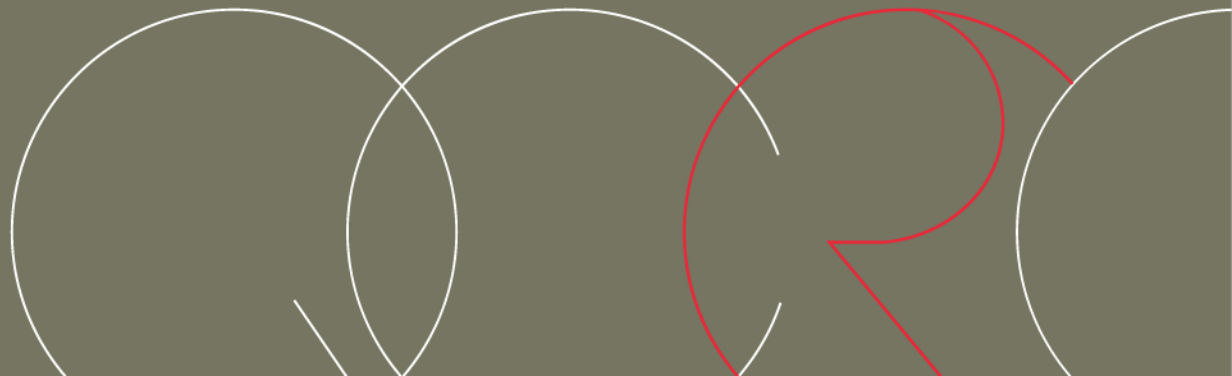# Demographic Surveys of Arab Annotators on CrowdFlower

Hamdy Mubarak, Kareem Darwish

{hmubarak, kdarwish}@qf.org.qa

Qatar Computing Research Institute, HBKU, Doha, Qatar

"Weaving Relations of Trust in Crowd Work: Transparency and Reputation across Platforms" workshop.

May 22, 2016. Hannover, Germany

Overview

- Motivation and Goal
- Related Work
- Survey Settings
- Survey Results
- Cross Survey Agreement
- Conclusions

Motivation and Goal

- Crowdsourcing (CS) is the process of segmenting a complex task into smaller units of work (Human Intelligence Tasks, or HIT's) and distributing them to be done by a large number of online workers (annotators) at lower monetary and time costs compared to traditional employees

- CS has advantages in: cost, speed, flexibility, scalability, and diversity

- Important issues for consideration are:
  - Worker demographic suitability: ex. language, age, education, etc.
  - Task complexity
  - Payout

- Goal: Examine such issues for CrowdFlower (CF) workers from Arab countries

Related Work

- Ipeirotis surveyed demographic information of 1,000 MTurk workers [Ipeirotis, 2010], including:

  - Gender, age, educational level, income level, marital status, number of HITs/week, and motivation

- CrowdFlower surveyed demographic information for 20,000 workers, including:

  – Country, age, number of children, education, ethnicity, gender, income, and marital status

  – (https://success.crowdower.com/hc/en-us/articles/202703345-Crowd-Demographics)

Survey Settings

- Two surveys of 500 CF workers each on June 22 and Aug. 4, 2015 (Survey 1 & 2) with "Language Capability" set to Arabic.

- Survey covers:
  - Age,
  - gender,
  - highest level of education,
  - foreign languages proficiency (English and French),
  - preferred pay rate for 1 minute of work,
  - country of origin, and
  - reason for working on CF.

- We ran the survey twice, because we suspected that some workers would contribute to both and hence we can determine answer consistency.
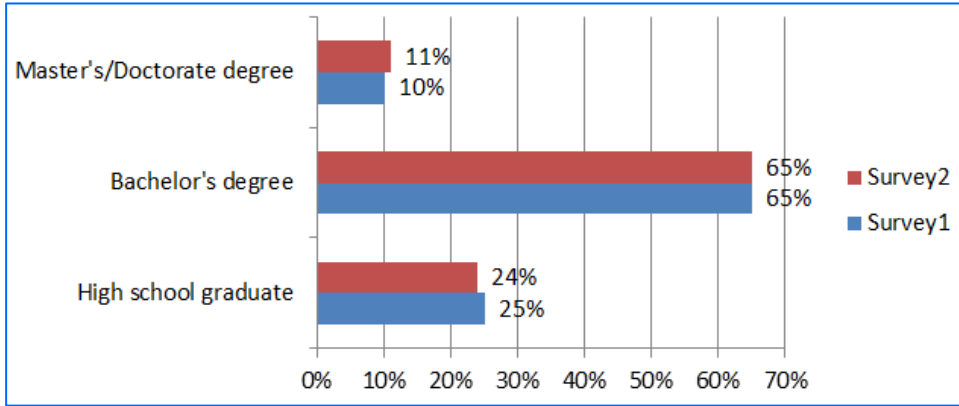
Survey Results

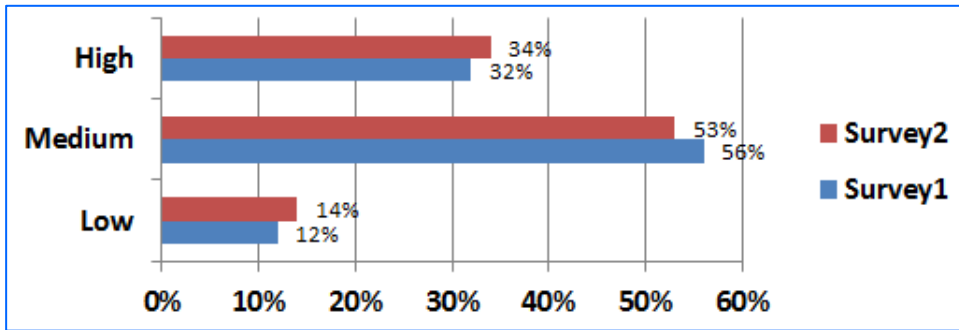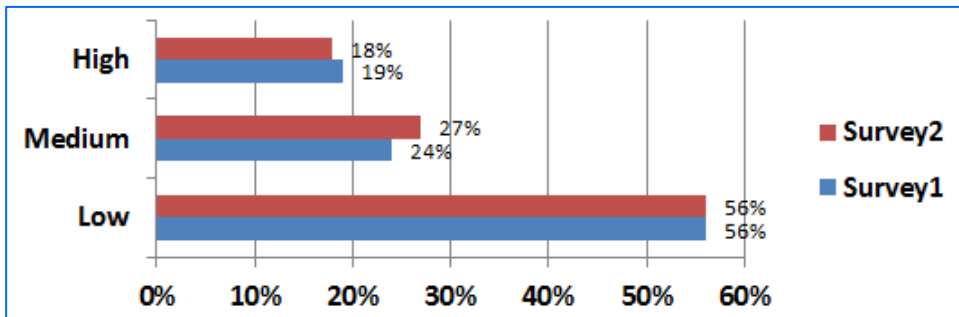Workers are mostly:

.. males (>75%)

.. aged 20-39 (>77%)

Qatar Computing Research Institute
Member of Qatar Foundation
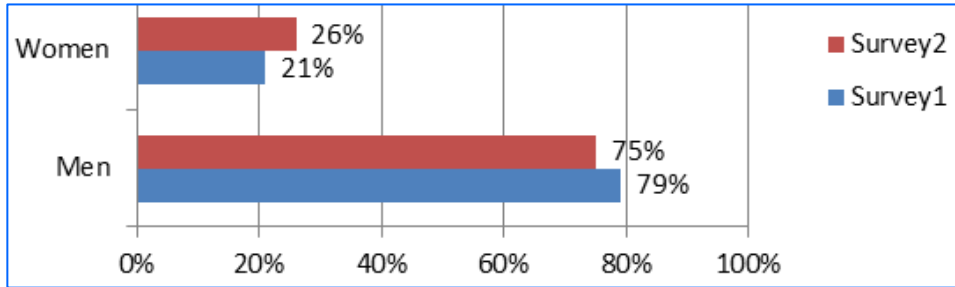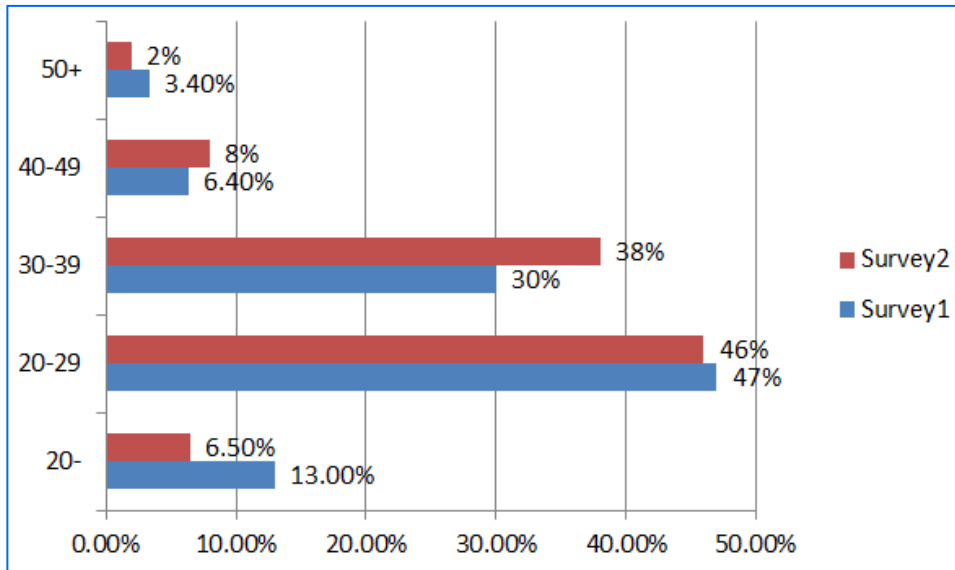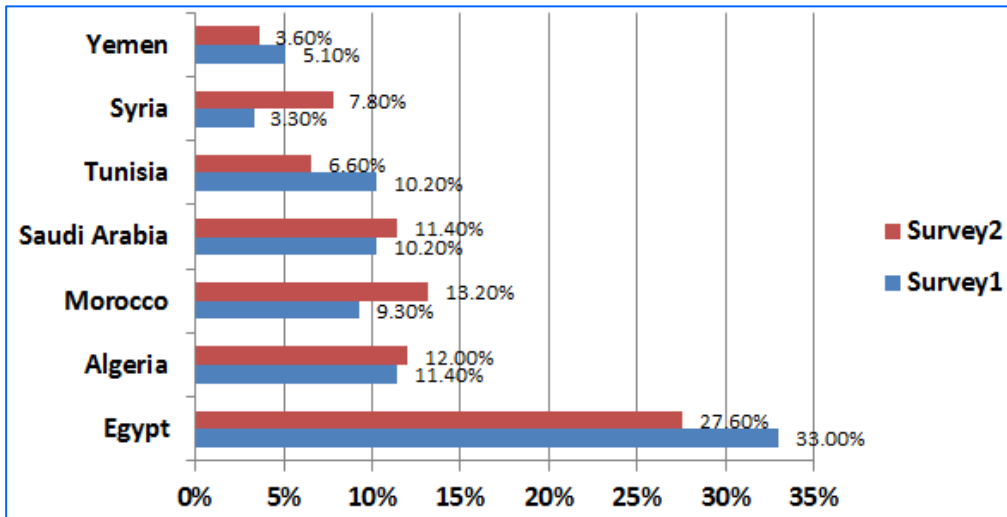
Survey Results

.. college educated (>75%)
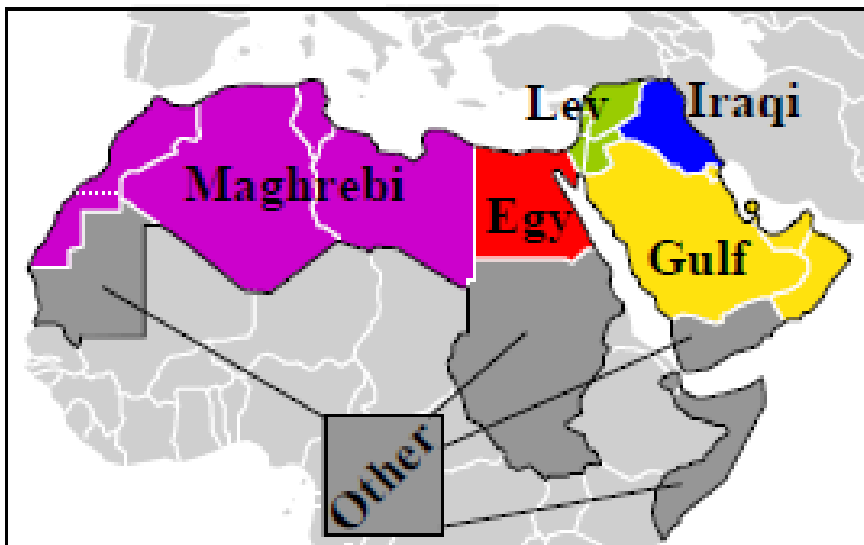
.. with medium/high English proficiency (>87%)

.. with low proficiency of French (>56%)

Survey Results

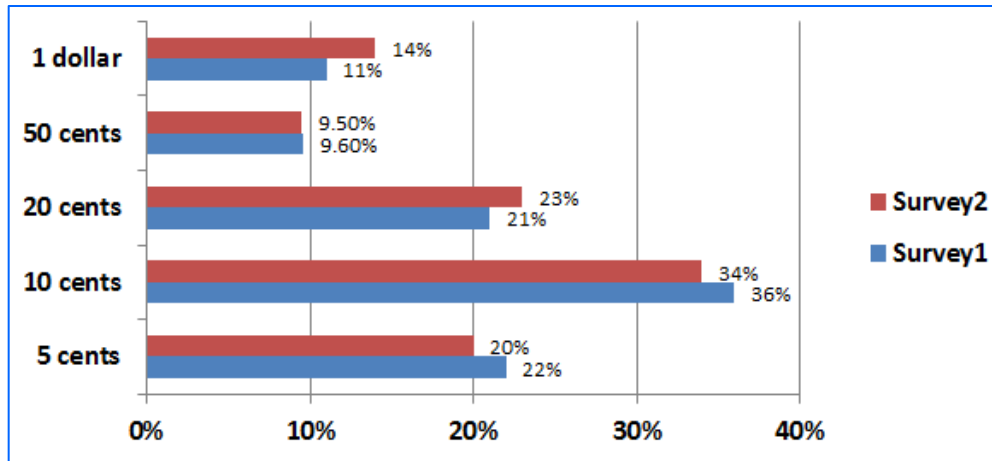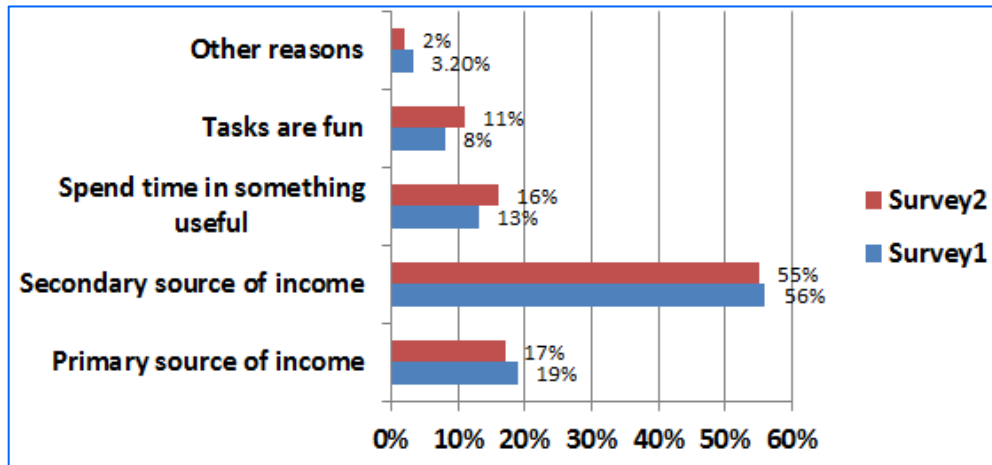- The country with the most number of workers is Egypt (30%), which is the most populous Arab country

- There are also workers from a variety of different countries that speak different dialects of Arabic (ex. Maghrebi (30%), Gulf (11%), Levantine (7%), and Yemeni (5%)).
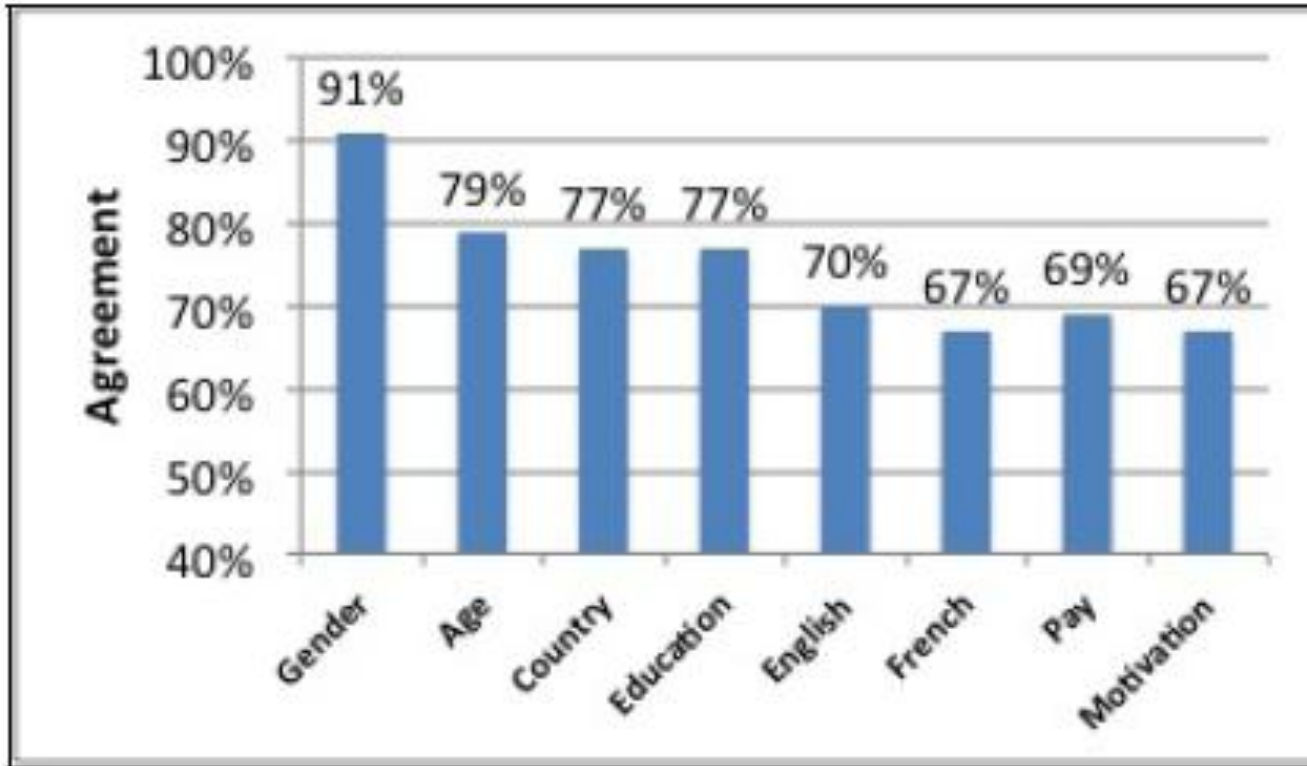
Survey Results

- Most of CF workers are welling to be paid 20 cents or less per minute for their work (~80%)

- Most of them work at CF as a secondary source of income (>55%)

Cross Survey Agreement

"Gender" and "Age" have highest agreement
"Pay Rate" and "Motivation" have lowest agreement

- One third of contributors participated in both surveys

- We used the agreement between survey items for common contributors as a measure of confidence

Conclusions

- We surveyed demographic information of Arab CF annotators, collected from two surveys carried out at time periods

- Considering the survey results can lead to enhanced annotation the quality

- From cross survey agreement, we can estimate that confidence of the quality of collected data to be around 80% on the average

Questions?