

Guidelines for Using CrowdTruth Framework

Oana Inel^{1,3}, Khalid Khamkham^{1,3}, Tatiana Cristea^{1,3}, Anca Dumitrache^{1,3},
Arne Rutjes², Jelle van der Ploeg², Lukasz Romaszko^{1,3}, Lora Aroyo¹, and
Robert-Jan Sips³

¹ VU University Amsterdam
k.khamkham@gmail.com, {anca.dumitrache,oana.inel,lora.aroyo}@vu.nl,
tatiana.cristea@student.vu.nl, lukasz.romaszko@gmail.com

² IBM Services Center Benelux, The Netherlands
arne.rutjesISC@nl.ibm.com, j.van.der.ploegISC@nl.ibm.com

³ CAS Benelux, IBM Netherlands
Robert-Jan.Sips@nl.ibm.com

1 Login in CrowdTruth

The CrowdTruth services are available at <http://crowdtruth.org>. Login is required in order to use the system. This allows us to record the actions of individual users and display this subsequently as part of the data statistics. In this way, each content unit in CrowdTruth can be searched by the user who ingested it in the system. To create a new account on CrowdTruth use this link: <http://crowdtruth.org/user/register> and invitation code "demo".

2 CrowdTruth Code

The open source CrowdTruth framework is available for download at <https://github.com/laroyo/CrowdTruth>.

3 CrowdTruth Modules

The CrowdTruth framework has three main interaction screens, i.e. **Media**, **Jobs and Workers**. You can access them at anytime from the top menu. For each of these screens two views are provided, a default **GraphView** with analytics of the current selection, and a corresponding **ListView**. You can switch between views by clicking the button in the top right corner.

In each of those views you can perform a number of search and filtering tasks to get to right selection of MediaUnits, Workers or Jobs you are looking for. Use the **filters** on the top of each column (in the ListView) and on the top of the overall pie charts (in the GraphView). Use also the **Specific Filters** button on the top left to apply filters specific to the select DocumentType only.

You can manipulate your view by selecting only the columns (properties of the data) you would like to see, i.e. use **Open All Columns** button dropdown to make your selection.

The **Actions** button is context dependent and provides different options for different type of Media, Jobs or Worker. Shared across all those data types are the actions for **save selection**, **export**. The MediaUnit **pre-processing** workflows are available as an option under this button.

The **Analytics GraphView** is organised as follows:

- **Top Pie Charts**, positioned at the top of the page, give an overview of the distribution of the properties in the current selection. For example, in Jobs, it will show the distribution of users executed jobs, the distribution of jobs across the different annotation tasks, different DocumentTypes, domains and formats. In MediaUnits, it will show the distribution of users created MediaUnits, the distribution of MediaUnits across the different annotation tasks, different DocumentTypes, domains and formats, as well as whether they have been used in a job or not.
- **Main Bar Chart**, positioned in the below the pie charts shows the individual items in the current selection, e.g. in Jobs each bar will be a job, in Media each bar will be a MediaUnit, in Workers, each bar will be a worker. Depending on the selection the corresponding properties and metrics are overlaid in this bar chart in order to compare your current selection to the metrics across collection, e.g. average number of annotations, average number of units per job, average MediaUnit clarity, etc. Detailed Pie Chart & Bar Char. These appear only if a selection is made in the Main Bar Chart. Both the pie chart and the bar chart are interactive, e.g. by selecting a portion of the pie chart the corresponding items in the bar char get highlighted. By selecting a bar in the bar chart, a popup with more information on CrowdTruth metrics is presented. As we Implement the triangle of reference, for each selection in the Main Bar Chart, we present three detailed Pie & Bar charts, i.e. one for the workers, one for the annotations and one for the MediaUnits.

Follow the flow in Figure 1 to experience the end-to-end workflow provided by the framework.

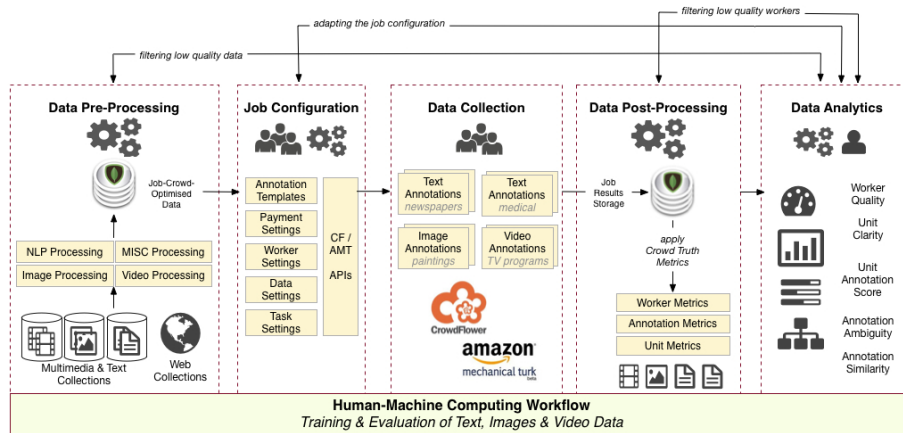


Fig. 1: CrowdTruth Overall Architecture

Follow the following workflows in some of the steps. For example, follow the data collection workflow depicted in Figure 3 to experience the configuration and

creation of a new job. Or follow the pre-processing workflows for text, images and videos depicted in Figure 2 for the specifics of each data modality.

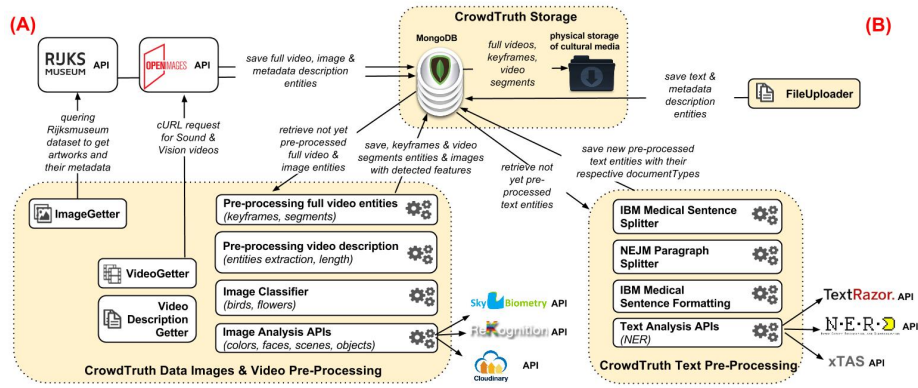


Fig. 2: CrowdTruth Pre-processing Workflows for Text, Images and Videos

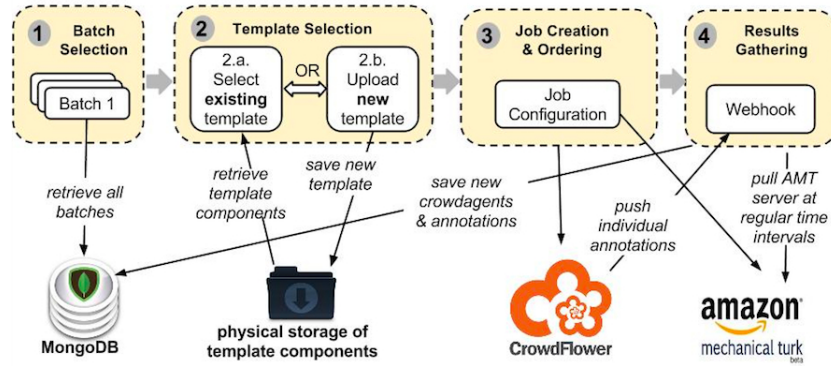


Fig. 3: CrowdTruth Data Collection Workflow

4 CrowdTruth Documentation

All the CrowdTruth documentation is available at <http://crowdtruth.org/info>. Previous published work can be found at <http://crowdtruth.org/papers>.