

This file contains details on the configuration.json parameters and the database specifications.

configuration.json parameters:

1. port_number: port information necessary to connect to the AWOCATo; default 8080.
2. accounts:
 - 2.1. username: specifies one new user and the respective username.
 - 2.2. password: specifies the password for the user in 2.1.
3. mongodb_settings_password-protected-db and mongodb_settings:
 - 3.1. url: url or IP address to the server hosting your MongoDB.
 - 3.2. port: Port under which your MongoDB is accessible; often: 27027.
 - 3.3. database: name of the MongoDB database.
 - 3.4. collection: name of the MongoDB collection.
 - 3.5. Items_per_query: number of items returned with each db query. Small numbers require more frequent queries, large ones occupy more cache; default: 10.
4. consolidation_mode: [true, false] whether the user wants to enable the consolidation mode or not; default false.
5. guideline_at_start: [true, false] to present the annotations guidelines each time the annotator(s) log-in; default true.
6. annotation_mode: [sentiment, sentiment-news, bi-classification, tri-classification, textual-annotation] to select one of the four supported modes; default sentiment.
7. mode_settings:
 - 7.1. standard_deviation_relevance: maximum accepted standard deviation for relevance annotations using a fine-grained scale. Texts exceeding this value are labeled for manual consolidation.
 - 7.2. polarity_threshold: threshold around 0 accepted as a “neutral” polarity rating.
 - 7.3. same_polarity_annotation: minimum required annotations agreeing on the polarity (this information is used for the automatic consolidation).
 - 7.4. data_specifications:
 - 7.4.1. fields_to_display_text: database field containing the text shown in box 5, figure 1.

- 7.4.2. `fields_to_display_title`: database field containing the text shown in box 4, figure 1.
- 7.4.3. `fields_to_display_entity`: database field containing the annotation target; only for the sentiment and sentiment-news modes (box 2, figure 1).
- 7.4.4. `fields_to_query`: all database fields required to populate the annotation interface (depending on the mode, 7.4.1, 7.4.2, 7.4.3)
- 7.4.5. `fields_to_store_in`: names of the database field(s) to store the annotations. The modes sentiment and sentiment-news require two fields.
- 7.4.6. `interface_settings`:
 - 7.4.6.1. `Idontknow_value`: The value assigned to an annotation when “I don’t know” is clicked.
- 7.4.7. `class_names`:
 - 7.4.7.1. `class_name`: Value/Text shown to the annotator on each class button (only used for the binary and ternary classification).
 - 7.4.7.2. `annotation_value`: value assigned to the annotation when the annotator clicks the corresponding button.
 - 7.4.7.3. `button_color_bg`: background color for the class name button specified in 7.4.6.1.
 - 7.4.7.4. `button_color_border`: border color for the class name button specified in 7.4.6.1.
- 7.4.8. `button_at_side`: [true, false] whether the class_name button appears below the annotation’s text or at the side; default False.

8. `consolidation_settings`:

- 8.1. `Minimum_annotations`: minimum number of annotations per text; default 3.
- 8.2. `standard_deviation`: allowed standard deviation for the annotations; default 0.21.
- 8.3. `accepted_annotators`: A list of user IDs considered for the consolidation (admin accounts can be excluded). The first specified username corresponds to 0, the second to 1, etc.

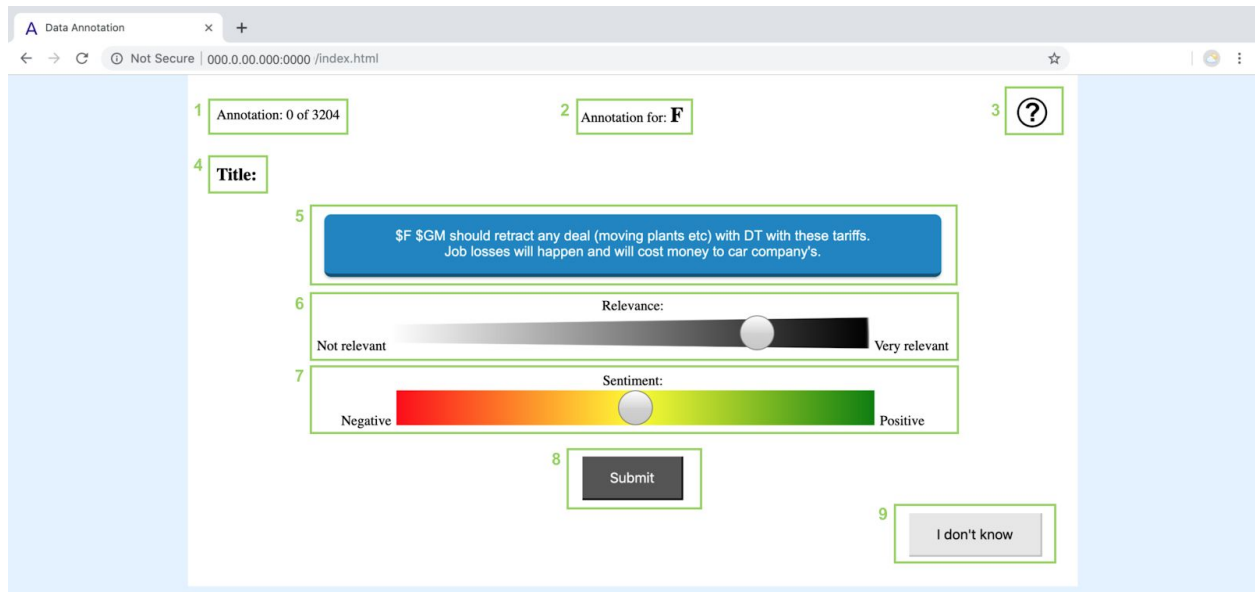


Figure 1 - AWOCAtO annotation interface using two fine-grained scales. Additional information can be found in the respective publication.

Database specifications:

According to the selected annotation mode, the data must contain the respective fields:

1. binary classification, ternary classification, and textual-annotation: "text" field with the text requiring annotation.
2. sentiment: "target" field specifying the target entity, and a "text" field as described above.
3. sentiment-news: "target", "text1", "text2", and "entity_spans"
 - 3.1. "target" specifies the targeted entity or entities
 - 3.2. "text1" is always shown to the annotator (e.g. a news title)
 - 3.3. "text2" has to be pre-processed to populate "entity_spans". For example, a news article can have multiple annotation targets; for each annotation target, the tool requires one data-tuple in the "entity_spans".
E.g.: [{target1, text2_1}, {target2, text2_2}]

Please note that text2_1 and text2_2 do not have to be the same string, thus, you can annotate different substrings for each annotation target. You find an example in sentiment-news.

- 3.4. “entity_spans” is a field of type list as specified above. While other field names can be personalized (amend the configuration.json accordingly), “entity_spans” is fixed.