

ISA-16 VoxML Track Annotation Guidelines

February 2020

Abstract

This task involves the annotation of images of objects and people participating in activities and events. The images are provided in the local subdirectory /VoxML-Track-images/. No captions are provided with the images. There is a vocabulary of objects, attributes, actions, and relations, with which annotators need to create an annotation of the image, as exhaustively as possible.

Guideline Specifications

The dataset consists of a set of images showing objects in common configurations or activities. For each image, we ask annotators to answer the following questions:

1. Provide a brief caption of the image.
2. What activities (if any) are the focus of the scene?
3. Identify the objects/entities in the scene (including people and animals).
4. Identify the major spatial and configurational relations between the objects/entities in the scene. Below are some sample relations:

left	right	in_front
behind	in	beside
on	above	containing
below	over	supporting
under	near	holding
touching	not_touching	

Annotations need not be restricted to this relation set but please adhere as closely to it as possible.

5. For each object, what activities **are shown** being performed with that object in the image?

- List up to 4.
6. For each object, what *other* activities **could** be performed with that object?
- List up to 4.
- (a) For each of those alternate activities, what circumstances need to be changed for this activity to be performed? Use the approximate format: *To <activity> <object>, <agent> **must** <action>*

Sample Annotation

Below is a sample image and possible annotations:



1. Provide a brief caption of the image.
A man and woman drinking together at a restaurant
2. What activities (if any) are the focus of the scene?
man drinking from glass
3. Identify the objects/entities in the scene (including people and animals).
man, woman, glasses/cups, bottles, sunglasses

4. Identify the major spatial and configurational relations between the objects/entities in the scene.
*man **holding** glass, woman **holding** cup, woman **beside** man, bottle **in front of** man, bottles **behind** woman, cup **in front of** woman, man **touching** glass, glass **containing** liquid, man **wearing** sunglasses, bottles **on** surface*
5. For each object, what activities **are shown** being performed with that object in the image?
 - glass:
man drinking from glass, man holding glass
 - cup:
woman holding cup
6. For each object, what *other* activities **could** be performed with that object?
 - cup:
woman drinking from cup
 - glass:
man setting down glass
 - bottle:
man/woman drinking from bottle
man/woman rolling bottle
- (a) For each of those alternate activities, what circumstances need to be changed for this activity to be performed?
 - *woman drinking from cup*
To drink from the cup, the woman must lift it to her mouth
 - *man setting down glass*
To set the glass down, the man must put it on a surface and release it
 - *man/woman drinking from bottle*
To drink from the bottle, the man/woman must open it and lift it to his/her mouth
 - *man/woman rolling bottle*
To roll the bottle, the man/woman must lay it down on a surface

Submission Requirements

Participants are required to use the annotation spreadsheet enclosed in this folder, **annotation_template.xlsx**, for entering all annotation information.

Annotators should attempt to follow the guidelines and stay within the provided vocabulary where possible, but are permitted to go outside if required. Participants are also required to submit a brief summary of and justification for any additions to the guideline specifications. We will provide feedback on submitted annotations. Deadline for submissions is **March 20, 2020**. We will select certain annotation submissions for spotlight presentations. These are to be given during the on-site portion of the workshop and will focus on the results of the annotation exercise. This will be followed by a discussion of the entire annotation effort, lessons learned (problems, gaps, suitability of results for ML, automated generation, etc), and the adequacy and expressiveness of the representation language.

Contact Information

For any questions please contact the track coordinator James Pustejovsky (jamesp@brandeis.edu) or Nikhil Krishnaswamy (nkrishna@brandeis.edu). As further background reading, see the LREC 2016 paper *VoxML: A Visualization Modeling Language*, by James Pustejovsky and Nikhil Krishnaswamy, enclosed in this directory or accessible here <https://arxiv.org/pdf/1610.01508.pdf>.