AIDA Annotation Guidelines:

Quality Control and Informative Mentions

V1.0

Linguistic Data Consortium
March 22, 2019

Table of Contents

0.0 Change log for these guidelines

- V1.0 is the first version of these guidelines to be distributed

1.0 Introduction to the AIDA Project

Active Interpretation of Disparate Alternatives (AIDA) is a project that focuses on multiple accounts of current events in multi-media data (including text, image, video, and mixed media). The data and also the categories that we annotate will be relevant to specific topics within an overarching scenario, and will involve several languages. The goal of AIDA Salient Mention Annotation is to create a corpus with multiple accounts of the same current events in news and social media, and to annotate the salient events and relations (along with their arguments) in each document.

The first overarching scenario for the project is the area of *Ukraine-Russia Relations 2014-1015*, and we will focus on a number of specific topics within that scenario. The languages for this scenario are Ukrainian, Russian, and English. The second overarching scenario will be *Crisis in Venezuela: 2010 to Present*, and the languages for that scenario will be Spanish, English, and Russian. The principles of annotation will be the same for each scenario, but this version of AIDA annotation guidelines will focus on the images (and ontology and categories) for the Ukraine-Russia Relations scenario.

Prior to Quality Control (QC) and Informative Mentions annotation, the Salient Mention annotators will annotate documents as they relate to each topic, focusing on capturing one salient mention for each knowledge element in each document element. The Quality Control (QC) and Informative Mentions annotation task is primarily to perform a second quality control pass of annotation over the document, and to also capture additional informative mentions in the process.

2.0 Quality Control (QC) and Informative Mention Annotation

The Quality Control (QC) and Informative Mentions annotation task is primarily to perform a second quality control pass of annotation over the document, and to also capture additional informative mentions in the process.

QC is the main goal of this task. We want every mention and every knowledge element to be as correct as possible, and we want to capture every salient mention.

In addition to general overall QC, this task will label additional mentions that add information to the mentions that were tagged during salient mention annotation. *We are not doing exhaustive annotation for this task!* But we are trying to capture the additional informative mentions that are in the documents.

2.1 Quality Control (QC) Annotation

The primary goal of this annotation task is to provide quality control (QC) for the Salient Mention annotation task that precedes this one.

> **Guiding Principle**: Apply the same standards and judgments that are described in the Salient Mention annotation guidelines.

Add missing annotation
In particular, focus on things that are salient to the topic but which were not annotated in the Salient Mention annotation pass. In particular, salient information about relation/event arguments should be added when that information is also salient.

Example:
- Present annotation: "shooting"
  - Type: LDC_evt_008 Conflict.Attack.FirearmAttack
  - Template: <arg1> attacked <arg2> using <arg3> at <arg4> place
  - Event argument roles:
    - Attacker: riot police snipers
      - Type: LDC_ent_106 PER.Combatant.Sniper
    - Target: Euromaidan protesters
      - Type: LDC_ent_126 PER.Protester.Unspecified
    - Place: Maidan
      - Type: LDC_ent_047 FAC.Structure.Plaza
    - When: 2014-02-20

- If not present, should be added: Affiliation relations (1) "riot police snipers" and (2) "Euromaidan protesters"
  - (1) Type: LDC_rel_035 OrganizationAffiliation.EmploymentMembership.Employment
  - Template: <arg1> is employed by <arg2>
  - Relation argument roles:
    - Employee: riot police snipers
      - Type: LDC_ent_106 PER.Combatant.Sniper
    - PlaceOfEmployment: riot police

- Type: LDC_ent_089
  ORG.Government.LawEnforcementAgency
  - When: Unknown

- (2) Type: LDC_rel_014
  GeneralAffiliation.MemberOriginReligionEthnicity.Unspecified
- Template: \<arg1\> is affiliated with \<arg2\>
- Relation argument roles:
  - Person: Euromaidan protesters
    - Type: LDC_ent_126 PER.Protester.Unspecified
  - EntityOrFiller: Euromaidan
    - Type: LDC_ent_133 SID.Ideological.Ideological
  - When: Unknown

Correct incorrect annotation

- Incorrect entity/filler extents
  - Example:
    - "Ukraine's security chief"
      - Correct: [Ukraine]'s security chief
      - Incorrect: [Ukraine's] security chief
    - "the president of Russia"
      - Correct: [the president of the United States]$_{NOM}$
      - Incorrect: the [president]$_{NOM}$ of the United States

- Mention level
  - Example:
    - "Berkut riot police"
      - Correct: [Berkut]$_{NAM}$ riot police
      - Correct: [Berkut riot police]$_{NOM}$
      - Incorrect: [Berkut riot police]$_{\text{NAM}}$
    - "The maidan"
      - Correct: The [maidan]$_{NAM}$
      - Incorrect: [The maidan]$_{\text{NOM}}$

- Type/subtype/subsubtype errors
  - Example:
    - "a corner of the square the police had captured in battles that began two days earlier"
      - Correct: FAC.Structure.**Plaza**
      - Incorrect: FAC.Structure.~~Unspecified~~
    - "The Maidan"
      - Correct: **FAC.Structure.Plaza**
      - Incorrect: ~~LOC.Position.Unspecified~~

- An explosion shown in the picture track of a video (see below)
    - ○ Correct: Conflict.Attack.**Bombing**
    - ○ Incorrect: Conflict.Attack.~~FirearmAttack~~

- Role label
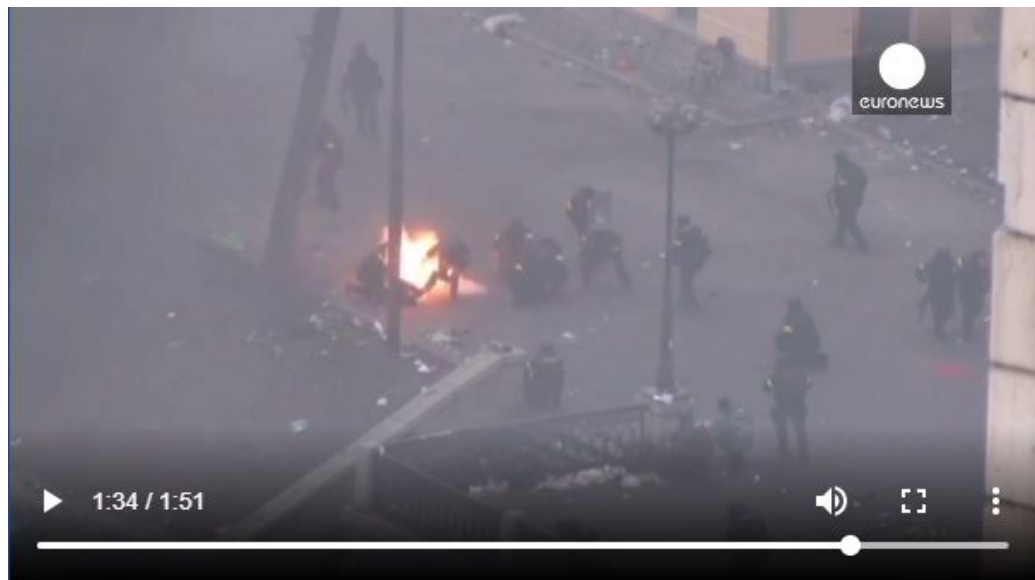   - Example: An image showing an explosion amidst police officers

- Correct:
    - ○ Type: LDC_evt_007 Conflict.Attack.Bombing
    - ○ Template: <arg1> attacked <arg2> using <arg3> at <arg4> place
    - ○ Event argument roles:
        - **Target**: police officers
            - Type: LDC_ent_111 PER.Police.Unspecified
        - Instrument: explosive
            - Type: LDC_ent_161 WEA.Bomb.Unspecified
        - Place: Maidan
            - Type: LDC_ent_047 FAC.Structure.Plaza
        - When: 2014-02-20

- Incorrect:
    - ○ Type: LDC_evt_007 Conflict.Attack.Bombing
    - ○ Template: <arg1> attacked <arg2> using <arg3> at <arg4> place
    - ○ Event argument roles:
        - ~~Attacker~~: police officers

- ● Type: LDC_ent_111 PER.Police.Unspecified
  - ■ ~~Target: protesters~~
    - ● Type: LDC_ent_126 PER.Protester.Unspecified
  - ■ Instrument: explosive
    - ● Type: LDC_ent_161 WEA.Bomb.Unspecified
  - ■ Place: Maidan
    - ● Type: LDC_ent_047 FAC.Structure.Plaza
  - ■ When: 2014-02-20

- Video/audio mentions have reasonable start/ends
  - Example:

- ● A single explosive device exploding in the sound track of a video
  - ○ Reasonable:
    - ■ Role: Instrument
    - ■ **Provenance**: 94.245285 : 95.338499
  - ○ Unreasonable:
    - ■ Role: Instrument
    - ■ Provenance: ~~0 : 111.966621~~

- Viktor Yanukovych describing his role in the Maidan violence in a press conference
  - Reasonable:
    - Event: Contact.PublicStatementInPerson.Unspecified
    - **Provenance**: 2.565931 : 76.356684
  - Unreasonable:
    - Event: Contact.PublicStatementInPerson.Unspecified
    - Provenance: ~~0 : 0.887594~~

- Temporal annotation
  - Example:
    - Correct: 20 Feb **2014**
    - Incorrect: 20 Feb _____ (year unspecified)

    - Correct: **Start Unknown**
    - Incorrect: ~~Started On XX-XXX-XXXX~~

Remove annotation that is not salient to the topic

In the following example, there is a potential contact event, but this additional mention does not provide information that is salient to the topic.

Example:

- But one pro-Ukraine teacher [said]~~Contact.Discussion.Meet~~ she received death threats after refusing to let rebels use her school as a polling station

In the following example, there is a potential relation, but it has effectively been annotated within an already-existing event.

Example:
- Ukrainian Interior Ministry has released new footage of riot police snipers in Kiev shooting at Euromaidan protesters on February 20.
  - Correct:
    - Type: LDC_evt_008 Conflict.Attack.FirearmAttack
    - Template: <arg1> attacked <arg2> using <arg3> at <arg4> place
    - Event argument roles:
      - Attacker: riot police snipers
        - Type: LDC_ent_111 PER.Police.Unspecified
      - **Place: Kiev**
        - Type: LDC_ent_057 GPE.UrbanAreaCity
      - Etc.

  - Incorrect:
    - (1) Type: LDC_evt_008 Conflict.Attack.FirearmAttack
    - Template: <arg1> attacked <arg2> using <arg3> at <arg4> place
    - Event argument roles:
      - Attacker: riot police snipers
        - Type: LDC_ent_111 PER.Police.Unspecified
      - Place: Kiev
        - Type: LDC_ent_057 GPE.UrbanAreaCity
      - Etc.

    - ~~(2) Type: LDC_rel_051 Physical.LocatedNear.Unspecified~~
    - ~~Template: <arg1> is located in or near <arg2>~~
    - ~~Relation argument roles:~~
      - ~~EntityOrFiller: riot police snipers~~
        - ~~Type: LDC_ent_111 PER.Police.Unspecified~~
      - ~~Place: Kiev~~
        - ~~Type: LDC_ent_057 GPE.UrbanAreaCity~~

> **Guiding Principle**: Trust the the judgment of the previous annotator. Unless you strongly believe the annotation is incorrect, give the previous annotator the benefit of the doubt and default to **not** changing the existing annotation.

2.2 Informative Mention Annotation

Do the QC annotation first! After QC is finished, then look for additional informative mentions. (So that you don't look for informative mentions of things that get deleted in QC…)

Salient Mention (and QC) annotation will have already captured the closest argument mentions. For this task, in addition to QC annotation of the full document, we will also tag additional mentions to capture additional information.

Definition: An *informative mention* is a mention of an entity/filler, relation or event that provides additional or more detailed information about an already tagged entity/filler, relation or event.

Tag additional mentions in each document element if they add either
- More information
- New information

Be sure to tag one mention of the most informative version in the document (if it was not already tagged).

Some types of informative mentions are
- A complete name (in text or sound)
  - Or a more complete name, or a named mention if only non-named mentions were previously tagged
- A nominal description that includes new or different information as compared to the closest mention that was tagged in Salient Mention Annotation
- A different version of the name, including different spellings (or misspellings) of the name, nicknames, acronyms, etc.
- The same entity in other media elements, even if there is no event/relation mention in the other media elements and only the entity/filler appears
- A more identifiable image of an entity in a video or image -- for example, a frame that shows a person's face

For example, if a person is mentioned as the argument of an event in a text doc element, and there are pictures of that person in several images and videos also in the document -- **all** of those mentions in the different images should be tagged as additional mentions, and at least one of the mentions in the video picture track (and audio track, if the mention appears there) should be tagged as additional mentions.

If in the course of adding informative mentions, you spot a salient event/relation that was missed during QC, then it's OK to add that as well, as a new mention.

Things you don't need to annotate for this task:
- Don't try to capture all of the pronouns for each mention -- we don't need to tag those

- Also, you don't need to annotate every nominal (or every named) mention -- tag the most informative mention you can find in the document element, and then tag mentions that add information of some kind

You only need to tag additional mentions for each **entity** one time. If the same entity participates in more than one event or relation, you only need to add the informative mentions to one of those entity/filler slots. (KB linking in the next stage of annotation will coreference entities and link all of these mentions together.)

Do try to capture all of the additionally informative mentions -- but remember that you should only tag additional mentions that add information. *And remember that we are not doing exhaustive annotation for this task!* We want to capture all of the **uniquely** informative mentions.

Example: *The new head of Ukraines **SBU** intelligence service Valentyn Nalivaichenko told reporters that his predecessor at **SBI**, Aleksandr Yakimenko oversaw the carnage in the Maidan on the orders of ex-president Yanukovych*

- Closest salient mention annotation of the ORG affiliated with Yakimenko:
  - SBI
- Additional informative mention:
  - SBU

Note that "SBI" in the above example is a typo for "SBU", but because SBI is the closest mention of the ORG argument for the affiliation relation with Yakimenko, that is the one that was tagged for Salient Mention Annotation, even though it is a typo. We capture SBU in informative mention annotation as a different spelling of the name of this entity (even though this "different" spelling is in fact the correct spelling).

Example: *The Alfa members were armed with AK-47s and sniper rifles, including Blaser hunting rifles.*

- Closest salient mention annotation of the weapons owned by Alfa members:
  - AK-47s
  - sniper rifles
- Additional informative mention:
  - Blaser hunting rifles

Note that in this example, you will also need to add a relation if it has not be tagged already, to capture "Blaser" as the Make of the hunting rifles.

2.2.1 Informative Mentions of Events and Relations

You also have the option of tagging additional informative mentions of Events and Relations, if you find a mention that adds information. Additional informative mentions of Events or Relations include things like a trigger/provenance that adds information.

Example: "gunned down" as an informative mention trigger for a previously tagged "shooting" event

Note that because in the original Salient Mention tagging of event/relation mentions, (1) arguments are drawn from the entire doc element, (2) salient event/relation mentions are already tagged in every doc element, and (3) event/relation mentions with conflicting information or conflicting arguments should already be tagged as separate mentions, we expect that the only *additional informative mentions* that will be taggable for events/relations are mentions that have different or more informative triggers or provenance.