Annotation guidelines for the debate.org dataset

**Goal:** Split debate posts into argumentative segments addressing different aspects of the topics.

Examples for aspects for selected topics are listed below.

Examples for popular controversial topics and aspects

* Climate change
* Gay marriage
* Student loan crisis
* Universal healthcare
* Euthanasia
* Marijuana legalization
* Israel
* Gun control
  + self-defense, self-reliance
  + guns don’t kill people – people kill people
  + rural life, farmers protecting themselves from wildlife
  + second amendment
  + shooting as sport
* Legalizing abortion
  + Health and safety of women
  + Body autonomy
  + Gender equality
  + Mental and physical consequences
  + Misusage for eugenics
  + Adoption as alternative
  + Life begins at conception
* School uniforms
  + Unity
  + Cost
  + Time saving
  + Bullying
  + Individuality

Annotation Process

**Requirements**

* GitHub: <https://github.com/anna---10/YEEDASeg.git>
* Start YEDDA\_Annotator.py in a terminal (requires Python 2.7 (e.g. in virtual environment))

**How to use YEDDA annotator**

* Add label: Place the cursor and press the key associated to the label you want to add
* Change label: Place the cursor within the existing label and press the key of the new label
* Remove label: Place cursor within an existing label and press Q

**Procedure**

1. Use Git PULL to get recent changes
2. Move the documents you want to annotate from *data* into your folder *AnnotatorXY*
3. Use Git ADD, COMMIT and PUSH to update these changes
4. Annotate the documents in your folder
   1. Add document label (see below for details)
   2. Add segment labels if document quality is medium or high (see below for details)
   3. Save the document
      1. If the same document is loaded again, the previous annotation will disappear
      2. Thus, open every document only once if you want to keep the annotations
5. Use Git ADD, COMMIT and PUSH again to upload your annotations

**3. Add document labels** (<low quality>, <medium quality>, and <high quality>)

To be placed at the very beginning of the document

1. <low quality> documents are documents that are not to be considered in the training process because they are either nonsense
   * *I’m a panda*
   * *Unicorns, Rainbows, Obama, and Randomness*
   * *It is probable that Adolf Hitler still lives*

or grammatically of very poor quality and thus not comprehensible

* + *Exercising is not the thing recess is and how could it raise obesity kids can just sneak food outside and just eat it JAYQUAN i know a lot of people that do that and at recess kids don't usually don't run they just walk around and do nothing and that's just it and its stupid to just go outside and just because you don't go outside dose not mean that kids cant go outside at home you can not do certain stuff at school recess like bring your phone.*
  + *First of all thier should be one way the human race should praise for or thier will think of themselfs as kings so god is real on a level but he is a sprit...*

No segment labels will be considered in combination with this document label

1. <high quality> documents show a well-structured argumentation and good grammar and spelling. Arguments are precise and supported by evidence. The text stands out by being good enough to be used in a school essay.
2. Most documents however will match the <medium quality> document label and are identified by being neither of low quality nor of high quality. E.g. they do make some points, but claims are not always supported by evidence, grammar and spelling could be better, they wander off the subject too much, etc. (examples are given on the following pages)

**4. Add segment labels** (<aspectXY>, <conclusion>, <outside of argumentation>)

Are to be placed in front of the starting word of the segment

* A segment is based on sentences and thus must contain at least one sentence! All text must be labeled with a segment label (in between the aspect and conclusion labels must be outside labels)

1. <aspectXY> labeled segments are mostly argumentative and can be distinguished from segments with other aspect labels by addressing a different aspect of the topic (see Example [4]). However, the same aspect label might be used for different segments addressing the same aspect.
   * A segment must show a stance towards the topic (see Example [5])
   * The statement must be comprehensible without the context of the discussion (apart from the topic)
   * If sensible, the segments should be chosen along line breaks and the pregiven structure (previous point) (see Example [1])
   * If there is something like “Point 1”, “1)”, “Argument A.” used by the authors to structure the argumentation, include this within the <aspectXY> segment, don’t label it as outside even if there is a little bit of outside talk like in this case:
     + *B. So what. The leaders in a Geniocracy would be incredibly selfish…*
     + Be consistent with this within a document
   * Also, if sensible, try to include transitional words like “First”, “Second”, “Next” within one segment as they indicate a start of a new point (see Example [3])
   * Go over your annotation at least once again to decide on the aspect labels
   * Look at the examples in the demotext directory and compare your annotation with the suggested one (X\_explanation.txt.ann)
2. The <conclusion> label is to be chosen for sentences concluding the whole document. Zero or one <conclusion> labels are allowed per document. Conclusions summarize the arguments an thus either appear at the beginning or the end of an argumentation. n
   * *If humans can change their definition of marriage as they please and there is no reason NOT to change it, why should we keep our laws about marriage the same?*
   * *Every able-bodied citizen in the US should attain at least 1 year of military training before the age of 24.*
3. The <outside of argumentation> label labels e.g. URLs, sentences addressing the opponent or the voters like
   * *I actually thought round 3 was the last round for some odd reason and treated it as such. I don't think there really is much to say here without bringing in new evidence so I'll just go over a few important points.*
   * *you still have not responded to my points about the poor earlier. food, health care, perhaps shelter. please refer to my anaologies and points in my first post.*

mere definitions like

* + *gentrification-the change in the character of a traditionally working-class area following an influx of new middle-class residents.*

or segments describing the background or stating an opinion but not being argumentative like

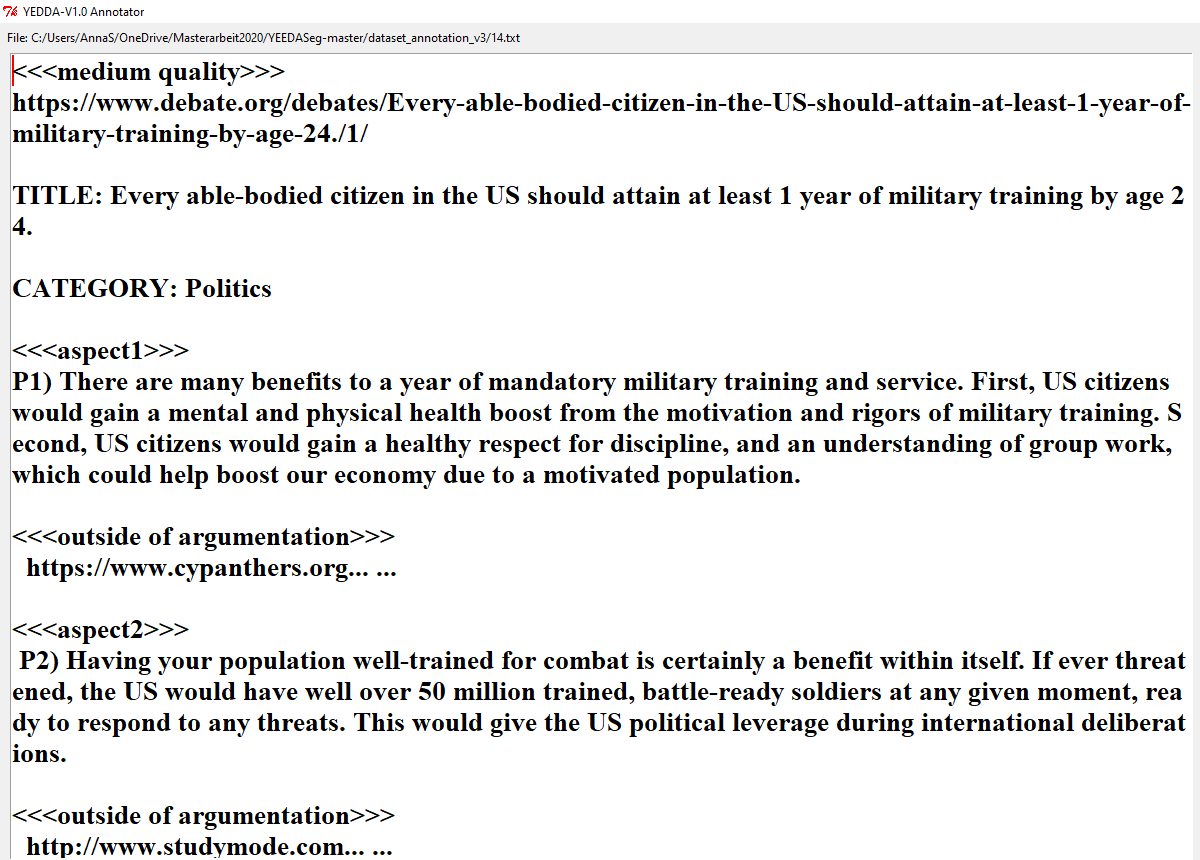
* + *With all the major shootings that have taken place recently, in the US and abroad, there has been an increased call for gun control in the United States. I assert that this is not only unconstitutional, but also would do little to nothing to combat mass gun violence.*

and statements not making a point towards the topic (see Example [2]).

Annotation examples

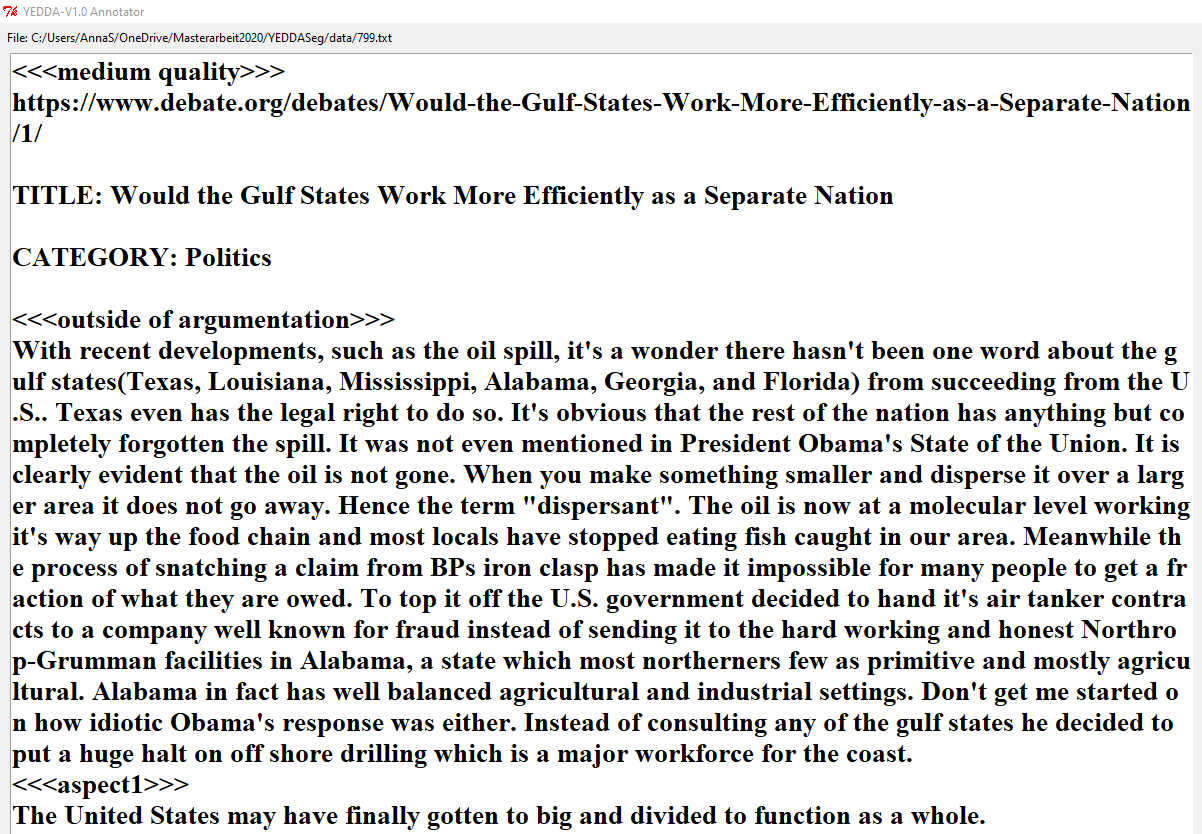
**Example [1]:**

Segments chosen along the document structure.

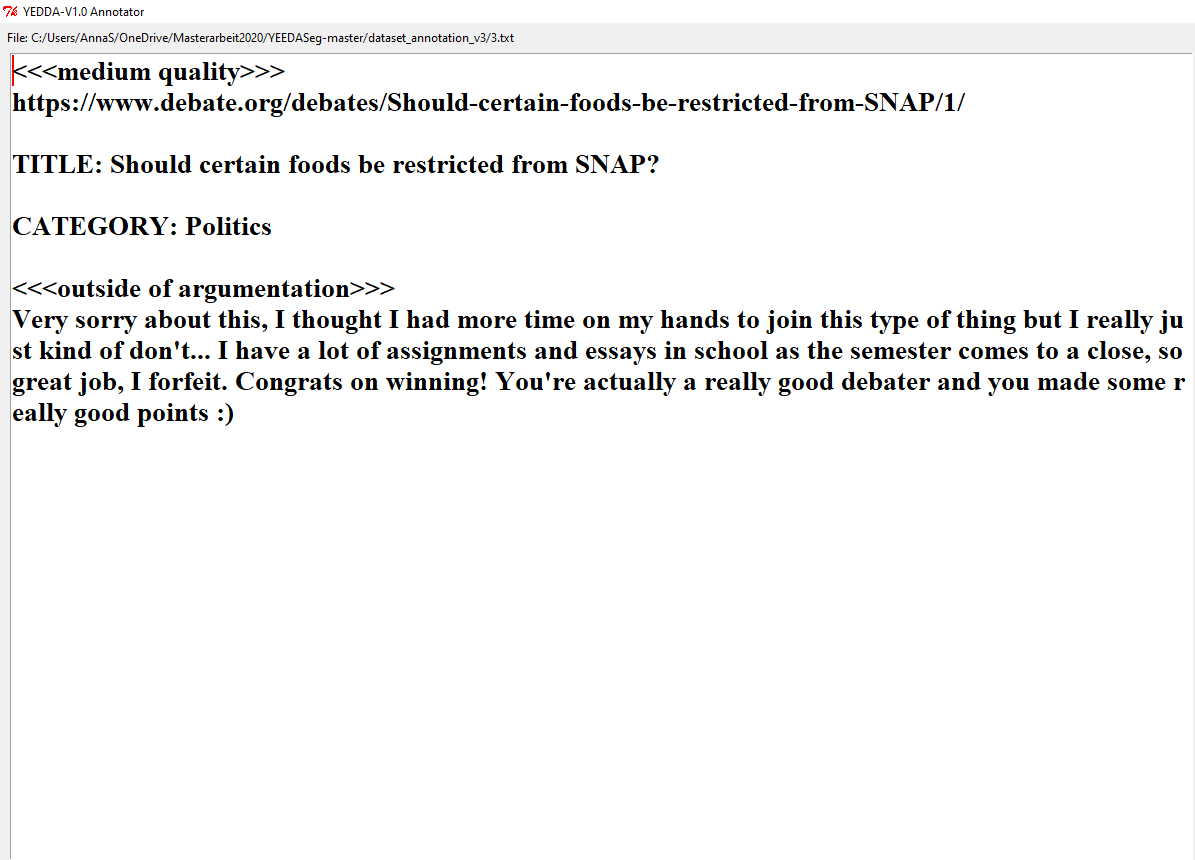


**Example [2]:**

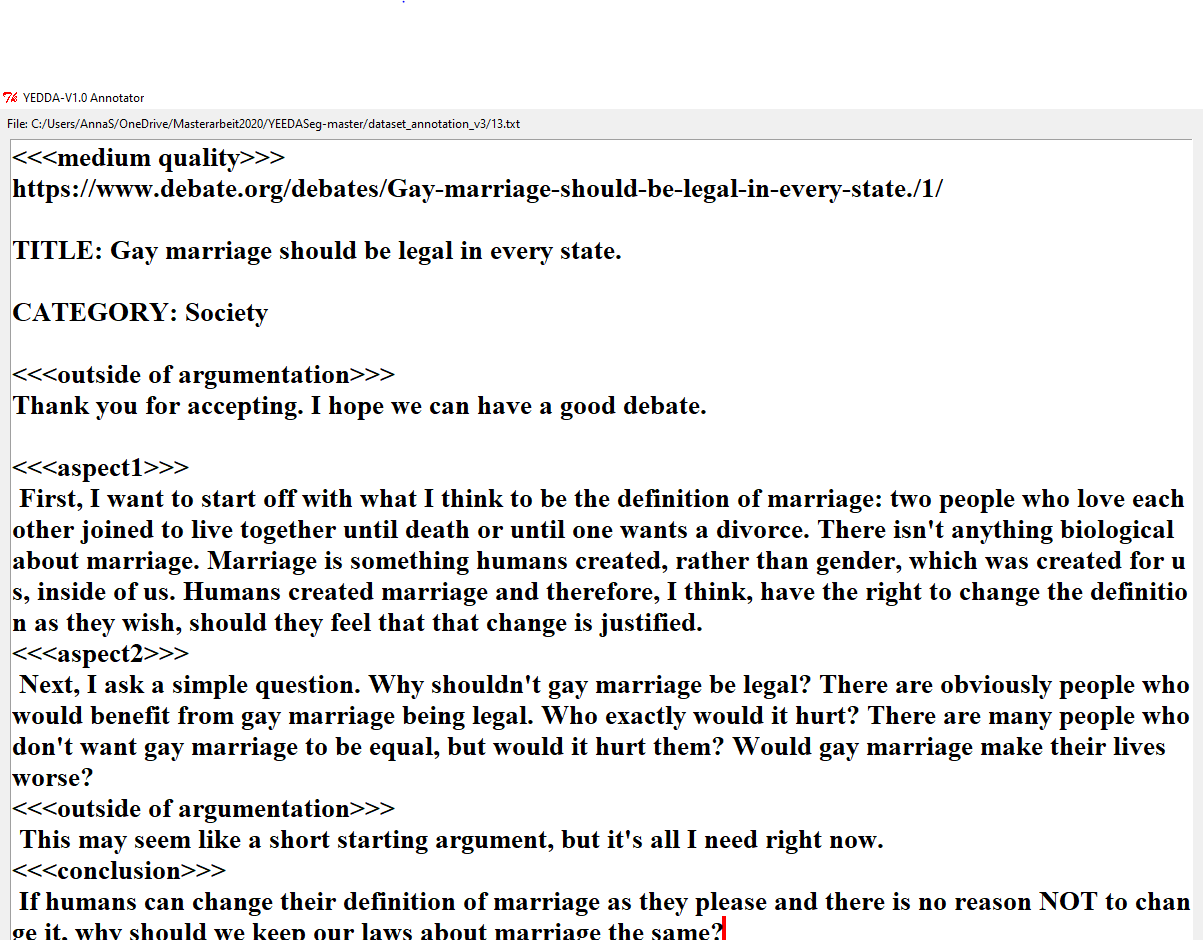
The following text is neither nonsense nor written poorly but it just makes one point which can be considered towards the topic. Thus, the document is labeled as <medium quality> and most of the text as <outside of argumentation>.



Similar situation for the following example: No point is made towards the topic and thus there is only one segment labeled as <outside of argumentation>.

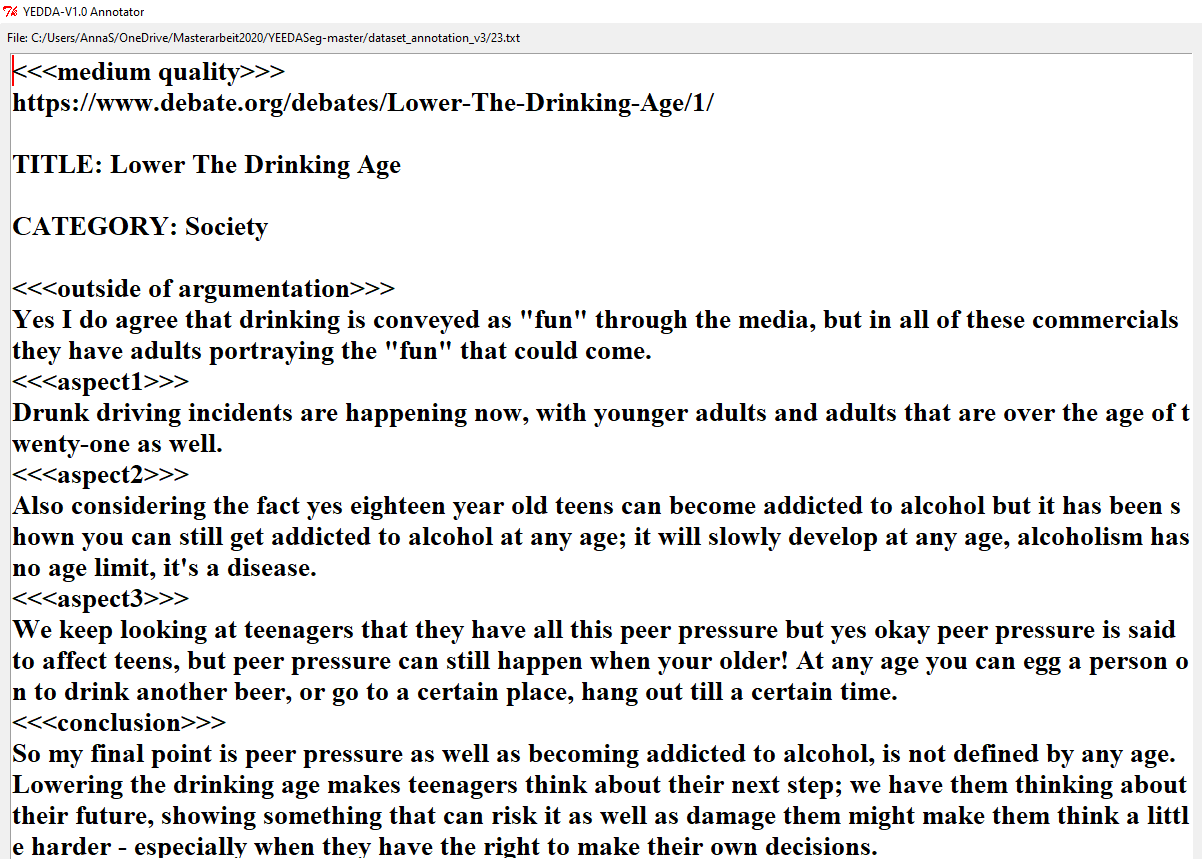


**Example [3]**

“Next, I ask a simple question.” could be considered outside the argumentation but is chosen to be inside to include the transitional word “Next” which indicates that a new point is made.

**Example [4]**

Text without line breaks and only a few transitional word is segmented by aspects.



Aspect1: Drunk driving

Aspect2: Alcoholism/ addiction

Aspect3: Peer pressure