

Args.me corpus cleansing evaluation: Annotator guidelines

1 Instructions

Argument corpora based on debate portals consist of human created argumentative text spans. Sometimes these text spans contain *information not relevant* to the argument proposed by the author. This irrelevant information can be of different nature. Some possible examples for such **irrelevant** sentences are spam, salutations, insults, meta-comments on the debate, expressions of gratitude or similar.

We are developing a semi supervised approach to automatically detect those sentences in order to remove them from the corpus since they do not hold important information regarding the argument. In order to evaluate the performance of the approach we need to analyse the generated results. By manually annotating the following dataset you will help us assess the approach's efficiency.

You will be presented with a list of sentences from the args.me corpus. Your task is to determine whether each individual sentence contains information related to an argument or has no value to the regarding argument.

Sentences are considered as **relevant** if they are connected to the argument in any way.

We consider a sentence as **irrelevant** if it does not support/present the (argumentative) opinion or facts offered by the author. See above for example kinds of such irrelevant sentences

Each sentence will be presented individually and the related argument is not shown. The sentence should be classifiable nevertheless. Some sentences may seem to be in relation to each other, but each sentence should be judged individually. It is possible that **all** sentences **or non** are relevant to an argument.

No sentences are included in the purpose of intentionally misleading you. We will not provide you with examples in order to not create a bias.

If you do have questions regarding the task, do not hesitate to ask!