

Understanding of Each column Header in CSV :

Fields Name	Description
Unannotated_id	A unique id for each instance created by hash combining of "Text" and "Root Url"
Text	The text on which annotation will be done
Overall_Sentiment	Overall Sentiment paragraph belongs NU= Neutral SP= Strongly Positive WP= Weakly Positive SN= Strongly Negative WN = Weakly Negative
Topic_Wise_Sentiment	Pair of topics and their topic wise sentiment NU= Neutral SP= Strongly Positive WP= Weakly Positive SN= Strongly Negative WN = Weakly Negative
Domain	The domain names from where the data belongs to
Validation	True = Manual validation is done after completing annotation by annotator and linguistic/validator False = Manual Validation not done
Root Url	The URL from where the data extracted from
Publish_date	The date when the data was published
Parse_time	Data scraping time

Annotator-1	Annotator basic info with which tag he/she has given to this particular data
Annotator-2	Annotator basic info with which tag he/she has given to this particular data
Annotator-3	Annotator basic info with which tag he/she has given to this particular data
Linguistic-1	Validator/linguistic basic info with which tag he/she has given to this particular data

Data Cardinality:

Number of instances: 20000

avg length of sentences: 48.95

median length of sentences: 43.0

mode length of sentences: 35

max length of sentences: 207

min length of sentences: 12

Number of words: 979021

Number of unique words: 78673

Number of unique words after stem: 61321

overall_kappa_score: 0.94