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
Subjectivity	Factual Neutral Data which has no polarity
Polarity	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 is 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: 70000
avg length of sentences: 9.91
median length of sentences: 9.0
mode length of sentences: 7
max length of sentences: 87
min length of sentences: 3
Number of sentences have less than 2 words: 0
Number of sentences have greater than 30 words: 490
Number of words: 693418
Number of unique words: 67911
Number of unique words after stem: 52195
kappa score: 0.883