

GENERAL INSTRUCTIONS FOR SENTENCE VALIDATION

Checks and tasks:

1. The task will be provided to you in google sheets. New tabs in this sheet will be created and assigned with different validation tasks as the validator progresses.

[illegible]

2. Tasks are mainly categorized into three types:
 - a. Sentence Level Checks
 - b. Word Level Checks
 - c. Character Level Checks
3. In sentence level checks, we programmatically filter out suspicious sentences for the validators to check correct if needed. These checks include several steps:
 - a. Sentences with invalid characters (action: remove invalid characters)
 - b. Sentences with dots (action: dots are properly used in acronyms, otherwise remove unwanted dots)
 - c. Similar sentences (action: checking and removal of redundant sentences)
 - d. Sentences with matra error (action: checking and removal of matra errors)
 - e. Sentence pruning (action: sentences are under a specified limit of characters)
 - f. Sentences with suspected acronyms (action: correct acronyms with proper format)
4. In word level checks, we programmatically filter out words susceptible to errors. These checks include:
 - a. Checking of the acronyms
 - b. Words susceptible to spelling mistakes
 - c. Words with halant (in Devnagri only)
 - d. Compound word check i.e. mistakenly joined words
5. In character level checks we only filter out the invalid characters. Invalid characters are as follows:
 - a. Characters except language's alphabets

- b. All numerals, even the language's script numerals.
- c. All punctuation except full stop(.), comma(,) and question mark(?)

Some Standard Guidelines:

1. The corrected sentences should be grammatically correct.
2. Make sure there are no typo errors in the corrected sentences.
3. You should ensure that there are no add digits in the sentences and should write the numbers in words only:
 - a. 132 will be written as one hundred thirty two
 - b. 1.32 as one point three two
 - c. ₹100 as rupees hundred
 - d. 90% as ninety percent
4. You should not add any special symbols (@#\$%^&*(){}[]+=|\/"'). In case the original sentence has any such characters, please do not use such characters in the translated/ corrected sentences column.
6. You should add only question marks (?), commas (,) and dots (.) and you should add dots only for the acronyms. You should also not add purna viram at the end of sentence

NotAllowed	Allowed
ATM	ए. टी. एम्
A.T.M	ए. टी. एम्
A.T.M.	ए. टी. एम्
ए. टी. एम् .	

OPTIONS GIVEN

- a. CORRECT (sentence has no issues at all and you can leave the Col E cell blank)
- b. INVALID SPL CHARS (if any special characters used. You have to also correct and write in col E)
- c. INCORRECT LANGUAGE (if sentence is not in your dialect, correct and write in col E)
- d. ACRONYMS (if dots are missing between acronyms (correct and write in col E)
- e. INCOMPLETE (try to complete the sentence, else you can just mark as incomplete and leave the Col E cell blank)
- f. CONTROVERSIAL (Need not correct the sentence, just mark as controversial)
- g. IDK: not sure of the sentence