

# VISUALLY GROUNDED PARAPHRASE EXTRACTION

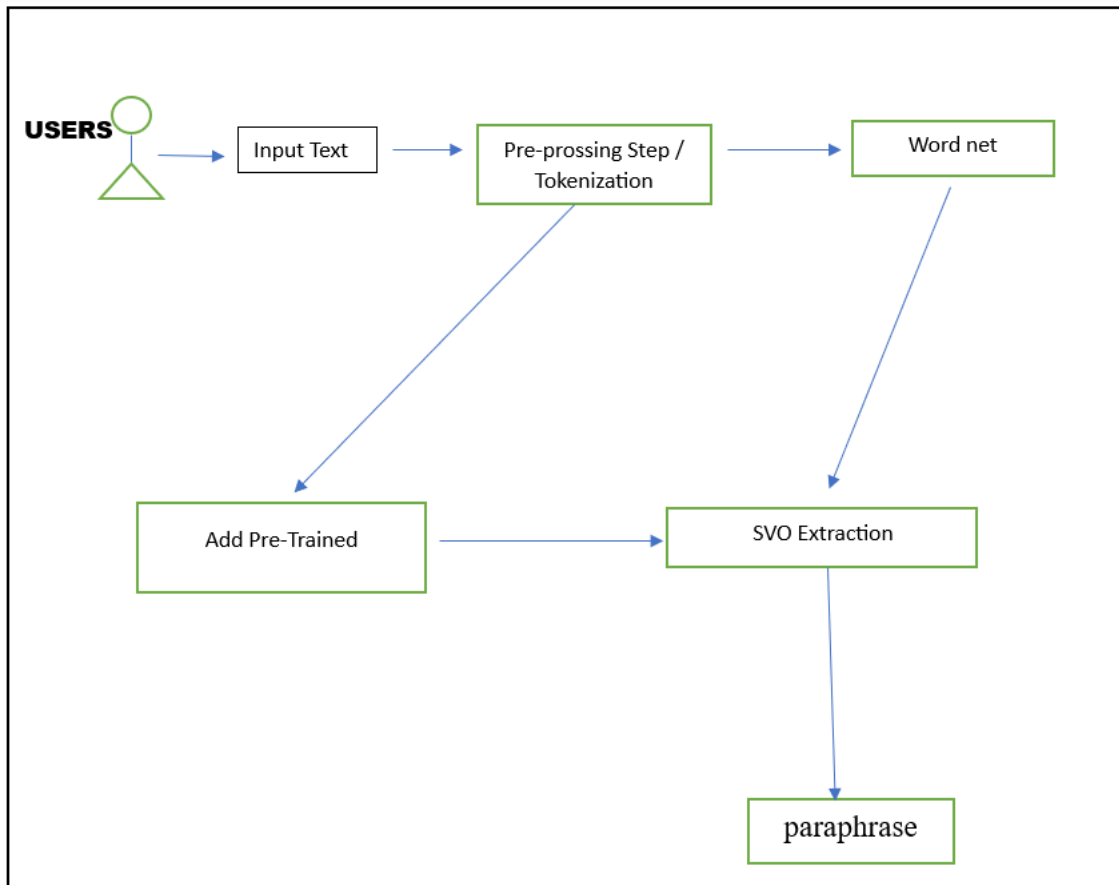


Centurion  
UNIVERSITY

**Group No - 07**

**SUPRIT-RANJAN**

**210101120217**



**User** - First step of paraphrase is that user. It may be he/she. User is defined as the person who will use this for his/her purpose.

**Input Data** - User can give any input that he/she wants to paraphrase it. It means that According to user need, he/she can generate paraphrase of that specific sentences and they can clearly understand after using it.

**Pre-processing Step** - after the giving input, next process is pre-processing Step. In this step, we have to pre-process the text that the sentences are Clean or not. If it is not clean, then it can clean by using two methods, 1. Wordnet and 2. Tokenizer ( it may be sent tokenizer or Word tokenizer) . If it is clean, there is no need of pre-processing Step.

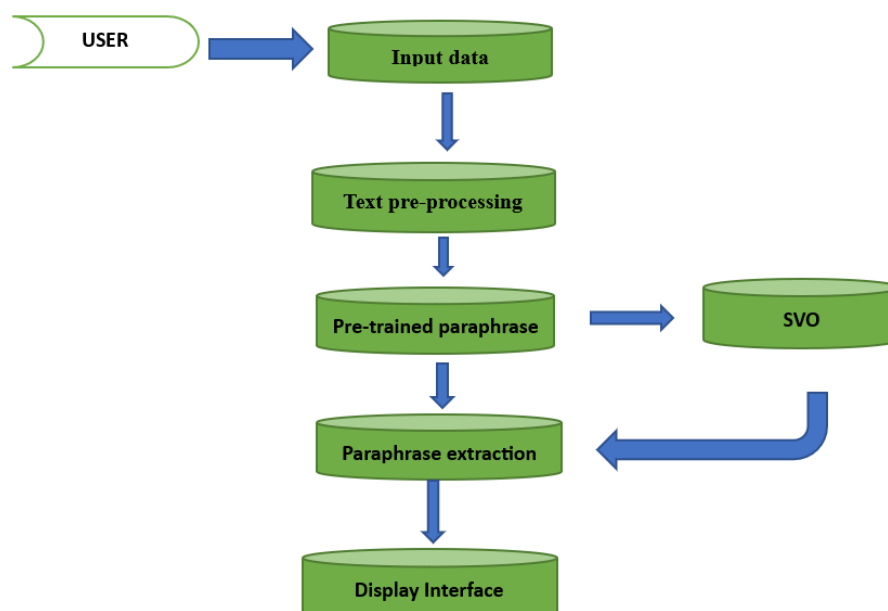
**Add pre train model** - then we have to add pre train model in it. some Pre - Trained model

( humarin/chatgpt\_paraphraser\_on\_T5\_base) we need for better understanding.

**paraphrase** - Add pre-trained model, we have to use T5 model and in this model we have to modify something, means in this model, we have to define some function i.e function paraphrase and when we define the function paraphrase, then next step is that we have to arrange the output of paraphrase in SVO arrangement. By getting this SVO arrangement output we have to add this algorithm. By using SVO arrangement, we have getting the output by a specific manner.

when we add function paraphrase and SVI arrangement in this pre-trained model, and our new model is ready to work. After process this all, our output will generate according to this functions.

DFD-



## Explanation of DFD

**1. Input data :-** We will take any sentence or text as input.

**2. Text pre-processing :-** It is process of cleaning and preparing raw text data for analysis .

**(i) Tokenization :-** Splitting text into individual word or tokens .

**(ii) Punctuation Removal:** Removing punctuation marks such as commas, periods, and exclamation points can make the text cleaner and easier to process.

**3. Pre-trained paraphrase model :-** Now, We will pretrained paraphrase model in this We have used "humarin/chatgpt\_paraphraser\_on\_T5\_base" and use of this model is a variant of GPT model for the specefic task and paraphrasing , Basically it train large amount of text data to generate paraphrased versions of text.

**4. Paraphrase extraction :-** Paraphrase extraction is the process of extracting pairs of sentences or phrases from text that have similar or equivalent meanings.

**5. Display Interface :-** Now , whatever paraphrase we have generated that will display.