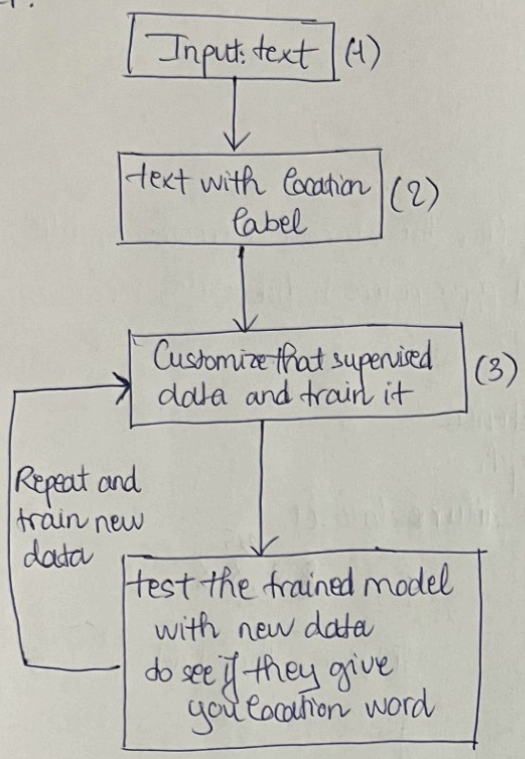
Ner:



The interface would look like:

Input: any text format (e.g. article)

Output: all the words meaning location and ~~the~~ a sentence or the surrounding words of that location words.

Example:

~~Ipe~~ Text input: " I like cat. I love NY and I want that one day I will go to Oxford University."

Output: - Location : NY → Country: United States.

- A sentence having the word: "I love NY and I want that one day I will go to Oxford University.

(1). Text pre-processing: check inside folder NER: folders PDF, WordDoc, excel, html. Right now my codes make all the format input to give the output as string text but the order of the words my be random. Just try 1 pdf / article and use pdf/html function and see the text output, you will understand what I meant.

(2) Standardize the location or pre-process the location-related words; "NY" should be "New York" (city) or "United States" (country). This will help when you input all the text, it might recognize the non-standardized location words and still know that the word is location word. (Check my "standardize_location.py")

(3) Check the "NER.py" and the youtube link I sent you to know how to use and customize NER or even train the data. (Notice: the youtube that guides you how to train NER right now is based on Rule-based NER which means you already known before the outlier cases might happen like NY is New York. But in the future if we haven't met these special cases, I still want our algorithm/model to learn heuristically not rigid rule-based. So be flexible and creative about this.