Process and Method:

1. Read the untagged file.

2. Divide the file into single line.

3. Using recursion:

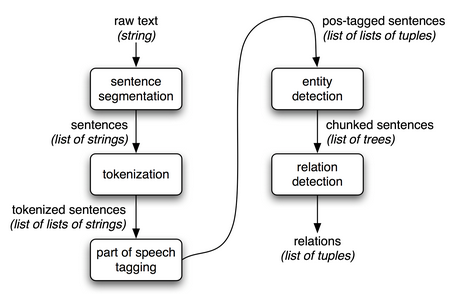
3.1 For date, time: make a collection of months and days and a function to decide if the number is between [1..31]. Then, split the line to words. To find if there is any word matches that collection or function. If there is, we should estimate the context nearby. eg. "1 May", if we find 1 is a number between [1..31], we should step to the next word to see if it is a month after. or "next Monday", if we find the key word "next", we should step to the next word to see if there is a word belongs to the months and days collection after "next".

or we can use regular expression to match the date and time.

3.2 For locations and person name: use gazetteer. eg. x `elem` (lines "persons.txt"). Of course, we should also take the nearby words into consideration.

3.3 For money: use key words and key symbol, such as million, thousand, hundred, $. format: $ + number + (million, thousand, hundred)

3.4 For organizations: begin with a capital letter or "Inc".



The figure is from: <http://www.nltk.org/book/ch07.html>

<https://github.com/Pispa339/HaskellNER>

When I did the NER coursework, I found a template on github (the link above). Based on it, I make some changes and add annotations for each function to make it more reasonable and understandable.