One of the most major forms of chunking in natural language processing is called "Named Entity Recognition." The idea is to have the machine immediately be able to pull out "entities" like people, places, things, locations, monetary figures, and more.

This can be a bit of a challenge, but NLTK is this built in for us. There are two major options with NLTK's named entity recognition: either recognize all named entities, or recognize named entities as their respective type, like people, places, locations, etc.

Here's an example:

import nltk

from nltk.corpus import state\_union

from nltk.tokenize import PunktSentenceTokenizer

train\_text = state\_union.raw("2005-GWBush.txt")

sample\_text = state\_union.raw("2006-GWBush.txt")

custom\_sent\_tokenizer = PunktSentenceTokenizer(train\_text)

tokenized = custom\_sent\_tokenizer.tokenize(sample\_text)

def process\_content():

try:

for i in tokenized[5:]:

words = nltk.word\_tokenize(i)

tagged = nltk.pos\_tag(words)

namedEnt = nltk.ne\_chunk(tagged, binary=True)

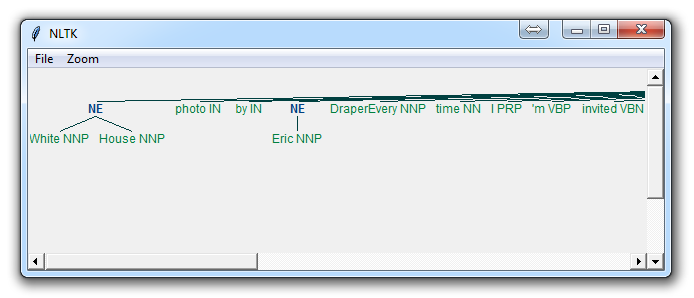
namedEnt.draw()

except Exception as e:

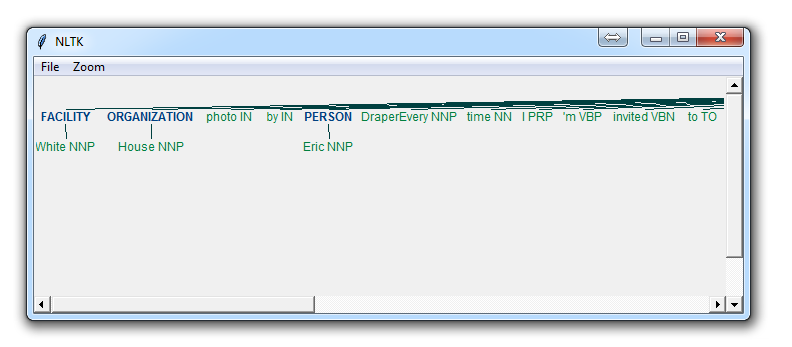
print(str(e))

process\_content()

Here, with the option of binary = True, this means either something is a named entity, or not. There will be no further detail. The result is:



If you set binary = False, then the result is:



Immediately, you can see a few things. When Binary is False, it picked up the same things, but wound up splitting up terms like White House into "White" and "House" as if they were different, whereas we could see in the binary = True option, the named entity recognition was correct to say White House was part of the same named entity.

Depending on your goals, you may use the binary option how you see fit. Here are the types of Named Entities that you can get if you have binary as false:

NE Type and Examples  
ORGANIZATION - Georgia-Pacific Corp., WHO  
PERSON - Eddy Bonte, President Obama  
LOCATION - Murray River, Mount Everest  
DATE - June, 2008-06-29  
TIME - two fifty a m, 1:30 p.m.  
MONEY - 175 million Canadian Dollars, GBP 10.40  
PERCENT - twenty pct, 18.75 %  
FACILITY - Washington Monument, Stonehenge  
GPE - South East Asia, Midlothian

Either way, you will probably find that you need to do a bit more work to get it just right, but this is pretty powerful right out of the box.