



DEMO PROJECT 2

PREPARED FOR

Dr. Sabine Bergler

PREPARED BY

Karandeep

40104845

TABLE OF CONTENTS

Introduction:	2
Purpose:	3
CFG (Context Free Grammar)/ Earley Parser:	3
Examples:	3
Pipeline:	4
Getting Started:	4
tokenizer():	4
Examples:	4
tokenizer() -> word_tokenizer:	9
Examples:	9
pos_tagger():	13
Examples:	13
measuredEntityDetection:	18
Examples:	18
dateRecognizer():	23
Examples:	23
dateParserCFG():	26
Examples:	26
named_entity():	32
Examples:	32

Introduction:

About NLTK - NLTK aka the Natural Language Toolkit, is a suite of open source Python modules, data sets, and tutorials supporting research and development in Natural Language Processing. It contains text processing libraries for tokenization, parsing, classification, stemming, tagging and semantic reasoning. It also includes graphical demonstrations and sample data sets as well as accompanied by a cook book and a book which explains the principles behind the underlying language processing tasks that NLTK supports.

Purpose:

Hidden information often lies deep within the boundaries of what we can perceive with our eyes and our ears. Some look to data for that purpose, and most of the time, data can tell us

more than we thought was imaginable. But sometimes data might not be clear cut enough to perform any sort of analytics. Language, tone, and sentence structure can explain a lot about how people are feeling, and can even be used to predict how people might feel about similar topics using a combination of the Natural Language Toolkit, a Python library used for analyzing text, and machine learning.

For this project I was to develop a **CFG** and a pipeline. The CFG It parses sentences into a parse tree that is a representation of **syntactic structure**, on which semantic interpretation can be based. It takes raw sentences either from inbuilt data or from the console. It uses pipelines `sent_tokenizer` and `word tokenizer` to split sentences into words and then draws sentences into a parse tree of the form **NP VP POS**.

The pipeline *Preprocess* parses by taking text using the NLTK corpus access commands. Analysis of data has been done in a way that is easily comprehensible. Overall goal of the project was to create a pipeline which filters data by using NLTK access commands and visualize them in a more informative way. NLTK is not perfect and lacks in some areas which is being explored below.

CFG (Context Free Grammar)/ Earley Parser:

Examples:

- Validation Data Sentence 1

OUTPUT:

```
-----> START <-----|
FOR SENTENCE----->      John ate an apple.

(S (NP (NNP John)) (VP (VBD ate) (NP (DT an) (NN apple))) (POS .))

-----> END <-----
```

- Validation Data Sentence 2

OUTPUT:

```
-----> START <-----
FOR SENTENCE----->      John ate the apple at the table.

(S
  (NP (NNP John))
  (VP
    (VBD ate)
    (NP (DT the) (NN apple))
    (PP (IN at) (NP (DT the) (NN table))))
  (POS .))

-----> END <-----
```

- Validation Data Sentence 3
OUTPUT:

```

-----> START <-----

FOR SENTENCE---->      On Monday, John ate the apple in the fridge.

(S
  (NP (PP (IN On) (NP (NNP Monday))) (POS ,) (NP (NNP John)))
  (VP
    (VBD ate)
    (NP (DT the) (NN apple))
    (PP (IN in) (NP (DT the) (NN fridge))))
  (POS .))

-----> END <-----
```

- Validation Data Sentence 4
OUTPUT:

```

-----> START <-----

FOR SENTENCE---->      On Monday, John ate the apple in his office.

(S
  (NP (PP (IN On) (NP (NNP Monday))) (POS ,) (NP (NNP John)))
  (VP
    (VBD ate)
    (NP (DT the) (NN apple))
    (PP (IN in) (NP (PRP his) (NN office))))
  (POS .))

-----> END <-----
```

- Validation Data Sentence 5
OUTPUT:

```

-----> START <-----

FOR SENTENCE---->      On Monday, John ate refrigerator apple in his office.

(S
  (NP (PP (IN On) (NP (NNP Monday))) (POS ,) (NP (NNP John)))
  (VP
    (VBD ate)
    (NP (NN refrigerator) (NN apple))
    (PP (IN in) (NP (PRP his) (NN office))))
  (POS .))

-----> END <-----

```

- Validation Data Sentence 6
OUTPUT:

```

-----> START <-----

FOR SENTENCE---->      Last week, on Monday, John finally took the apple from the fridge to his office.

(S
  (NP
    (NP (JJ Last) (NN week))
    (POS ,)
    (PP (IN on) (NP (NNP Monday)))
    (POS ,)
    (NP (NNP John))
    (ADVP (RB finally)))
  (VP
    (VBD took)
    (NP (DT the) (NN apple))
    (PP (IN from) (NP (DT the) (NN fridge)))
    (PP (IN to) (NP (PRP his) (NN office))))
  (POS .))

-----> END <-----

```

- Validation Data Sentence 7
OUTPUT:

```

FOR SENTENCE---->      Last Monday, John promised that he will put an apple in the fridge. He will eat it on Tuesday at his desk. It will be crunchy.

(S
  (NP (NP (JJ Last) (NNP Monday)) (POS ,) (NP (NNP John)))
  (VP
    (VBD promised)
    (SBAR
      (IN that)
      (S
        (NP (PRP he))
        (VP
          (MD will)
          (VP
            (VB put)
            (NP (DT an) (NN apple))
            (PP (IN in) (NP (DT the) (NN fridge)))))))
    (POS .))
  (S
    (NP (PRP He))
    (VP
      (MD will)
      (VP
        (VB eat)
        (NP (PRP it))
        (PP (IN on) (NP (NNP Tuesday)))
        (PP (IN at) (NP (PRP his) (NN desk))))
      (POS .))
    (S
      (NP (PRP It))
      (VP (MD will) (VP (VB be) (ADJP (JJ crunchy))))
      (POS .))
  )
)

-----> END <-----

```

- Validation Data Sentence 8

OUTPUT:

```

-----> START <-----
FOR SENTENCE----->      On Monday, September 17, 2018, John O'Malley promised his colleague Mary that he would put a replacement apple in the office fridge.
O'Malley intended to share it with her on Tuesday at his desk and anticipated that the crunchy treat would delight them both. But she was sick that day.

(S
(NP
  (PP
    (IN On)
    (NP
      (NP (NNP Monday))
      (POS ,)
      (NP (DATE (MONTH September) (DAY 17) (POS ,) (YEAR 2018))))
      (POS ,)
      (NP (NNP John) (NNP O))
      (POS ')))
      (NP (NNP Malley)))
  (VP
    (VBD promised)
    (NP (PRP his) (NN colleague) (NNP Mary))
    (SBAR
      (IN that)
      (S
        (NP (PRP he))
        (VP
          (MD would)
          (VP
            (VB put)
            (NP (DT a) (NN replacement) (NN apple))
            (NP
              (PP (IN in) (NP (DT the)))
              (NP (NN office) (NN fridge))))))
          (POS .))
        (S
          (NP (NP (NNP O)) (POS ')) (NP (NNP Malley)))
          (VP
            (VBD intended)
            (S
              (VP
                (IN to)
                (VP
                  (VB share)
                  (NP (PRP it))
                  (PP (IN with) (NP (PRP her)))
                  (PP (IN on) (NP (NNP Tuesday)))
                  (PP (IN at) (NP (PRP his) (NN desk))))
                (CC and)
                (VP
                  (VBD anticipated)
                  (SBAR
                    (IN that)
                    (S
                      (NP (DT the) (NN crunchy) (NN treat))
                      (VP
                        (MD would)
                        (VP (VB delight) (NP (PRP them)) (NP (DT both))))))
                      (POS .))
                    (S
                      (NP (CC But) (NP (PRP she)))
                      (VP (VBD was) (ADJP (JJ sick)) (NP (DT that) (NN day)))
                      (POS .))
                    (S

```

- Validation Data Sentence 9

OUTPUT:

-----> START <-----

FOR SENTENCE-----> Sue said that on Monday, September 17, 2018, John O'Malley promised his colleague Mary that he would put a replacement apple in the office fridge and that O'Malley intended to share it with her on Tuesday at his desk.

```
(S
  (NP (NNP Sue))
  (VP
    (VBD said)
    (SBAR
      (IN that)
      (S
        (NP
          (PP
            (IN on)
            (NP
              (NP
                (NP (NNP Monday))
                (POS ,)
                (NP
                  (DATE
                    (MONTH September)
                    (DAY 17)
                    (POS ,)
                    (YEAR 2018))))
                (POS ,)
                (NP (NNP John) (NNP O))
                (POS ')))
              (NP (NNP Malley)))
          (VP
            (VBD promised)
            (NP (PRP his) (NN colleague) (NNP Mary))
            (SBAR
              (IN that)
              (S
                (NP (PRP he))
                (VP
                  (MD would)
                  (VP
                    (VB put)
                    (NP (DT a) (NN replacement) (NN apple))
                    (NP
                      (PP (IN in) (NP (DT the)))
                      (NP (NN office) (NN fridge))))))))
                (CC and)
                (SBAR
                  (IN that)
                  (S
                    (NP (NP (NNP O)) (POS ')) (NP (NNP Malley)))
                    (VP
                      (VBD intended)
                      (SBAR
                        (IN to)
                        (S
                          (VP
                            (VB share)
                            (NP (PRP it))
                            (PP (IN with) (NP (PRP her)))
                            (PP (IN on) (NP (NNP Tuesday)))
                            (PP (IN at) (NP (PRP his) (NN desk))))))))
                          (POS .))
                    )
                  )
                )
              )
            )
          )
        )
      )
    )
  )
)
```

-----> END <-----

Pipeline:

Getting Started:

1. Open CMD (for Windows), Terminal (for MacOS).
2. Goto directory of the script.
3. Type in command "python3 preprocess.py"

tokenizer():

Examples:

- Example 1:
DATA -

AUSTN FEB ANNUAL M3 MONEY SUPPLY RISES 11.2 PCT
The annual growth of Australia's m3
money supply rose by 11.2 pct in the year ended February
compared with January's 10.7 pct, the Reserve Bank said.
This was down from 14.0 pct in February last year.
In February m3 rose by 0.6 pct compared with 0.8 in January
and a February 1986 rise of 0.1 pct
More

OUTPUT -

-----START-----

----> Tokened Sentences <----

SENTENCE ->

AUSTN FEB ANNUAL M3 MONEY SUPPLY RISES 11.2 PCT

SENTENCE ->

The annual growth of Australia's m3
money supply rose by 11.2 pct in the year ended February
compared with January's 10.7 pct, the Reserve Bank said.

SENTENCE ->

This was down from 14.0 pct in February last year.

SENTENCE ->

In February m3 rose by 0.6 pct compared with 0.8 in January
and a February 1986 rise of 0.1 pct.

SENTENCE ->

More

- Example 2:

DATA -

R.R. DONNELLY <DNY> TO SELL CABLE SYSTEM
R.R. Donnelly and Sons Co said it
reached an agreement with Adams-Russell Co Inc <AAR> to sell
Adams its Rockford, Ill., cable system for an undisclosed
price.

R.R. Donnelly said it expects the deal to close in about 120 days, subject to regulatory approvals and other conditions.

The cable system serves about 51,000 subscribers in the communities of Rockford, Loves Park, Machesney Park, Cherry Valley, Morristown and Winnebago County, Ill., R.R. Donnelly said.

OUTPUT -

----> Tokened Sentences <----

SENTENCE ->

R.R.

SENTENCE ->

DONNELLY <DNY> TO SELL CABLE SYSTEM

R.R.

SENTENCE ->

Donnelly and Sons Co said it

reached an agreement with Adams-Russell Co Inc <AAR> to sell Adams its Rockford, Ill., cable system for an undisclosed price.

SENTENCE ->

R.R.

SENTENCE ->

Donnelly said it expects the deal to close in about

120 days, subject to regulatory approvals and other conditions.

SENTENCE ->

The cable system serves about 51,000 subscribers in the

communities of Rockford, Loves Park, Machesney Park, Cherry Valley, Morristown and Winnebago County, Ill., R.R.

SENTENCE ->

Donnelly

said.

- Example 3:

DATA -

C.T.C. DEALER EXTENDS BID FOR CANADIAN TIRE

C.T.C. Dealer Holdings Ltd said it

extended its previously announced offer for 49 pct of <Canadian Tire Corp Ltd> to midnight on March 26.

CTC, a group of Canadian Tire dealers, which already owns 17 pct of Canadian Tire, is currently appealing in an Ontario court against a previously announced Ontario Securities Commission ruling blocking CTC's 272 mln dlr bid.

The blocked bid did not include non-voting shareholders, who hold 96 pct of Canadian Tire equity.

OUTPUT -

```
----> Tokened Sentences <----

SENTENCE ->
C.T.C.
SENTENCE ->
DEALER EXTENDS BID FOR CANADIAN TIRE
C.T.C.
SENTENCE ->
Dealer Holdings Ltd said it
  extended its previously announced offer for 49 pct of &lt;Canadian
  Tire Corp Ltd> to midnight on March 26.
SENTENCE ->
CTC, a group of Canadian Tire dealers, which already owns
  17 pct of Canadian Tire, is currently appealing in an Ontario
  court against a previously announced Ontario Securities
  Commission ruling blocking CTC's 272 mln dlr bid.
SENTENCE ->
The blocked bid did not include non-voting shareholders,
  who hold 96 pct of Canadian Tire equity.
```

- Example 4:

DATA -

N.Z. UNEMPLOYMENT RATE 3.9 PCT IN DECEMBER QUARTER

New Zealand's unemployment rate was 3.9 pct of the workforce in the quarter ended December 31, unchanged from a revised 3.9 pct (preliminary 3.8 pct) in the previous quarter but slightly above 3.8 pct in the year-earlier quarter, the Statistics Department said.

The Department, citing the Household Labour Force Survey, said in a statement the number of unemployed in October-December 1986 was 61,500 against 60,500 in the September quarter and 60,500 a year earlier.

OUTPUT -

```
----> Tokened Sentences <----

SENTENCE ->
N.Z.
SENTENCE ->
UNEMPLOYMENT RATE 3.9 PCT IN DECEMBER QUARTER
New Zealand's unemployment rate was
  3.9 pct of the workforce in the quarter ended December 31,
  unchanged from a revised 3.9 pct (preliminary 3.8 pct) in the
  previous quarter but slightly above 3.8 pct in the year-earlier
  quarter, the Statistics Department said.
SENTENCE ->
The Department, citing the Household Labour Force Survey,
  said in a statement the number of unemployed in October-
  December 1986 was 61,500 against 60,500 in the September
  quarter and 60,500 a year earlier.
```

- Example 5:

DATA -

M.A. HANNA CO <HNM> TO ACQUIRE <PMS CONSOLIDATED>
M.A. Hanna Co said it has reached
definitive agreement to purchase PMS Consolidated, a privately
owned manufacturer of colorants for the plastics industry.
M.A. Hanna said the purchase, which is subject to certain
conditions, is scheduled to close around the end of October.
Hanna announced last August that it was in negotiaitons to
purchase the company.
Headquartered in Somerset, N.J., PMS reported over 70 mln
dlrs in revenues in its last fiscal year, which ended May 1987.

OUTPUT -

```
-----START-----  
-----  
  
----> Tokened Sentences <----  
  
SENTENCE ->  
M.A.  
SENTENCE ->  
HANNA CO &lt;HNM> TO ACQUIRE &lt;PMS CONSOLIDATED>  
M.A.  
SENTENCE ->  
Hanna Co said it has reached  
    definitive agreement to purchase PMS Consolidated, a privately  
    owned manufacturer of colorants for the plastics industry.  
SENTENCE ->  
M.A.  
SENTENCE ->  
Hanna said the purchase, which is subject to certain  
    conditions, is scheduled to close around the end of October.  
SENTENCE ->  
Hanna announced last August that it was in negotiaitons to  
    purchase the company.  
SENTENCE ->  
Headquartered in Somerset, N.J., PMS reported over 70 mln  
    dlrs in revenues in its last fiscal year, which ended May 1987.
```

- Example 6:

DATA -

INDONESIA UNLIKELY TO IMPORT PHILIPPINES COPRA
Indonesia is unlikely to import copra
from the Philippines in 1987 after importing 30,000 tonnes in
1986, the U.S. Embassy's annual agriculture report said.
The report said the 31 pct devaluation of the Indonesian
rupiah, an increase in import duties on copra and increases in
the price of Philippines copra have reduced the margin between
prices in the two countries.
Indonesia's copra production is forecast at 1.32 mln tonnes
in calendar 1987, up from 1.30 mln tonnes in 1986.

OUTPUT -

-----> Tokened Sentences <-----

```
SENTENCE ->
INDONESIA UNLIKELY TO IMPORT PHILIPPINES COPRA
SENTENCE ->
  Indonesia is unlikely to import copra
  from the Philippines in 1987 after importing 30,000 tonnes in
  1986, the U.S. Embassy's annual agriculture report said.
SENTENCE ->
The report said the 31 pct devaluation of the Indonesian
  rupiah, an increase in import duties on copra and increases in
  the price of Philippines copra have reduced the margin between
  prices in the two countries.
SENTENCE ->
Indonesia's copra production is forecast at 1.32 mln tonnes
  in calendar 1987, up from 1.30 mln tonnes in 1986.
```

tokenizer() -> word_tokenizer:

Examples:

- Example 1:

DATA -

```
AUSTN FEB ANNUAL M3 MONEY SUPPLY RISES 11.2 PCT
The annual growth of Australia's m3
money supply rose by 11.2 pct in the year ended February
compared with January's 10.7 pct, the Reserve Bank said.
This was down from 14.0 pct in February last year.
In February m3 rose by 0.6 pct compared with 0.8 in January
and a February 1986 rise of 0.1 pct
More
```

OUTPUT-

-----> Tokened Words and Characters<-----

```
['AUSTN', 'FEB', 'ANNUAL', 'M3', 'MONEY', 'SUPPLY', 'RISES', '11.2', 'PCT',
'The', 'annual', 'growth', 'of', 'Australia', "'", 'm3', 'money', 'supply',
'rose', 'by', '11.2', 'pct', 'in', 'the', 'year', 'ended', 'February', 'com
pared', 'with', 'January', "'", '10.7', 'pct', ',', 'the', 'Reserve', 'Bank
', 'said', '.', 'This', 'was', 'down', 'from', '14.0', 'pct', 'in', 'Februar
y', 'last', 'year', '.', 'In', 'February', 'm3', 'rose', 'by', '0.6', 'pct',
'compared', 'with', '0.8', 'in', 'January', 'and', 'a', 'February', '1986',
'rise', 'of', '0.1', 'pct', '.', 'More']
```

-----> End of sentence token module <-----

- Example 2:

DATA -

R.R. DONNELLY & DNY> TO SELL CABLE SYSTEM

R.R. Donnelly and Sons Co said it reached an agreement with Adams-Russell Co Inc & AAR> to sell Adams its Rockford, Ill., cable system for an undisclosed price.

R.R. Donnelly said it expects the deal to close in about 120 days, subject to regulatory approvals and other conditions.

The cable system serves about 51,000 subscribers in the communities of Rockford, Loves Park, Machesney Park, Cherry Valley, Morristown and Winnebago County, Ill., R.R. Donnelly said.

OUTPUT -

----> Tokened Words and Characters<----

```
['R.R.', '.', 'DONNELLY', '&', 'lt', ';', 'DNY', '>', 'TO', 'SELL', 'CABLE',  
'SYSTEM', 'R.R.', '.', 'Donnelly', 'and', 'Sons', 'Co', 'said', 'it', 'reache  
d', 'an', 'agreement', 'with', 'Adams-Russell', 'Co', 'Inc', '&', 'lt', ';',  
'AAR', '>', 'to', 'sell', 'Adams', 'its', 'Rockford', ',', 'Ill.', ',', 'ca  
ble', 'system', 'for', 'an', 'undisclosed', 'price', '.', 'R.R.', '.', 'Donne  
lly', 'said', 'it', 'expects', 'the', 'deal', 'to', 'close', 'in', 'about',  
'120', 'days', ',', 'subject', 'to', 'regulatory', 'approvals', 'and', 'othe  
r', 'conditions', '.', 'The', 'cable', 'system', 'serves', 'about', '51,000',  
, 'subscribers', 'in', 'the', 'communities', 'of', 'Rockford', ',', 'Loves',  
'Park', ',', 'Machesney', 'Park', ',', 'Cherry', 'Valley', ',', 'Morristown',  
, 'and', 'Winnebago', 'County', ',', 'Ill.', ',', 'R.R.', '.', 'Donnelly', '  
said', '.']
```

----> End of sentence token module <----

- Example 3:

DATA -

C.T.C. DEALER EXTENDS BID FOR CANADIAN TIRE

C.T.C. Dealer Holdings Ltd said it extended its previously announced offer for 49 pct of & Canadian Tire Corp Ltd> to midnight on March 26.

CTC, a group of Canadian Tire dealers, which already owns 17 pct of Canadian Tire, is currently appealing in an Ontario court against a previously announced Ontario Securities Commission ruling blocking CTC's 272 mln dlr bid.

The blocked bid did not include non-voting shareholders, who hold 96 pct of Canadian Tire equity.

OUTPUT -

----> Tokened Words and Characters<----

```
['C.T.C', '.', 'DEALER', 'EXTENDS', 'BID', 'FOR', 'CANADIAN', 'TIRE', 'C.T.C',
',', '.', 'Dealer', 'Holdings', 'Ltd', 'said', 'it', 'extended', 'its', 'previ-
ously', 'announced', 'offer', 'for', '49', 'pct', 'of', '&', 'lt', ';', 'Can-
adian', 'Tire', 'Corp', 'Ltd', '>', 'to', 'midnight', 'on', 'March', '26', '
.', 'CTC', ',', 'a', 'group', 'of', 'Canadian', 'Tire', 'dealers', ',', 'whi-
ch', 'already', 'owns', '17', 'pct', 'of', 'Canadian', 'Tire', ',', 'is', 'c-
urrently', 'appealing', 'in', 'an', 'Ontario', 'court', 'against', 'a', 'pre-
viously', 'announced', 'Ontario', 'Securities', 'Commission', 'ruling', 'blo-
cking', 'CTC', '"s', '272', 'mln', 'dlr', 'bid', '.', 'The', 'blocked', 'bid',
', 'did', 'not', 'include', 'non-voting', 'shareholders', ',', 'who', 'hold',
', '96', 'pct', 'of', 'Canadian', 'Tire', 'equity', '.']
```

----> End of sentence token module <----

- Example 4:

DATA -

N.Z. UNEMPLOYMENT RATE 3.9 PCT IN DECEMBER QUARTER

New Zealand's unemployment rate was

3.9 pct of the workforce in the quarter ended December 31,
unchanged from a revised 3.9 pct (preliminary 3.8 pct) in the
previous quarter but slightly above 3.8 pct in the year-earlier
quarter, the Statistics Department said.

The Department, citing the Household Labour Force Survey,
said in a statement the number of unemployed in October-
December 1986 was 61,500 against 60,500 in the September

quarter and 60,500 a year earlier.

OUTPUT -

----> Tokened Words and Characters<----

```
['N.Z', '.', 'UNEMPLOYMENT', 'RATE', '3.9', 'PCT', 'IN', 'DECEMBER', 'QUARTE-
R', 'New', 'Zealand', '"s', 'unemployment', 'rate', 'was', '3.9', 'pct', 'of',
', 'the', 'workforce', 'in', 'the', 'quarter', 'ended', 'December', '31', ',',
', 'unchanged', 'from', 'a', 'revised', '3.9', 'pct', '(', 'preliminary', '3
.8', 'pct', ')', 'in', 'the', 'previous', 'quarter', 'but', 'slightly', 'abo-
ve', '3.8', 'pct', 'in', 'the', 'year-earlier', 'quarter', ',', 'the', 'Stat-
istics', 'Department', 'said', '.', 'The', 'Department', ',', 'citing', 'the',
', 'Household', 'Labour', 'Force', 'Survey', ',', 'said', 'in', 'a', 'statem-
ent', 'the', 'number', 'of', 'unemployed', 'in', 'October-', 'December', '19
86', 'was', '61,500', 'against', '60,500', 'in', 'the', 'September', 'quarte-
r', 'and', '60,500', 'a', 'year', 'earlier', '.']
```

----> End of sentence token module <----

- Example 5:

DATA -

M.A. HANNA CO <HNM> TO ACQUIRE <PMS CONSOLIDATED>

M.A. Hanna Co said it has reached definitive agreement to purchase PMS Consolidated, a privately owned manufacturer of colorants for the plastics industry.

M.A. Hanna said the purchase, which is subject to certain conditions, is scheduled to close around the end of October. Hanna announced last August that it was in negotiations to purchase the company.

Headquartered in Somerset, N.J., PMS reported over 70 mln dlrs in revenues in its last fiscal year, which ended May 1987.

OUTPUT -

----> Tokened Words and Characters<----

```
['M.A.', '.', 'HANNA', 'CO', '&', 'lt', ';', 'HNM', '>', 'TO', 'ACQUIRE', '&', 'lt', ';', 'PMS', 'CONSOLIDATED', '>', 'M.A.', '.', 'Hanna', 'Co', 'said', 'it', 'has', 'reached', 'definitive', 'agreement', 'to', 'purchase', 'PMS', 'Consolidated', ',', 'a', 'privately', 'owned', 'manufacturer', 'of', 'colorants', 'for', 'the', 'plastics', 'industry', '.', 'M.A.', '.', 'Hanna', 'said', 'the', 'purchase', ',', 'which', 'is', 'subject', 'to', 'certain', 'conditions', ',', 'is', 'scheduled', 'to', 'close', 'around', 'the', 'end', 'of', 'October', '.', 'Hanna', 'announced', 'last', 'August', 'that', 'it', 'was', 'in', 'negotiations', 'to', 'purchase', 'the', 'company', '.', 'Headquartered', 'in', 'Somerset', ',', 'N.J.', ',', 'PMS', 'reported', 'over', '70', 'mln', 'dlrs', 'in', 'revenues', 'in', 'its', 'last', 'fiscal', 'year', ',', 'which', 'ended', 'May', '1987', '.']
```

----> End of sentence token module <----

- Example 6:

DATA-

INDONESIA UNLIKELY TO IMPORT PHILIPPINES COPRA

Indonesia is unlikely to import copra from the Philippines in 1987 after importing 30,000 tonnes in 1986, the U.S. Embassy's annual agriculture report said.

The report said the 31 pct devaluation of the Indonesian rupiah, an increase in import duties on copra and increases in the price of Philippines copra have reduced the margin between prices in the two countries.

Indonesia's copra production is forecast at 1.32 mln tonnes in calendar 1987, up from 1.30 mln tonnes in 1986.

OUTPUT -

----> Tokened Words and Characters<----

```
['INDONESIA', 'UNLIKELY', 'TO', 'IMPORT', 'PHILIPPINES', 'COPRA', 'Indonesia', 'is', 'unlikely', 'to', 'import', 'copra', 'from', 'the', 'Philippines', 'in', '1987', 'after', 'importing', '30,000', 'tonnes', 'in', '1986', ',', 'the', 'U.S.', 'Embassy', "'s", 'annual', 'agriculture', 'report', 'said', '.', 'The', 'report', 'said', 'the', '31', 'pct', 'devaluation', 'of', 'the', 'Indonesian', 'rupiah', ',', 'an', 'increase', 'in', 'import', 'duties', 'on', 'copra', 'and', 'increases', 'in', 'the', 'price', 'of', 'Philippines', 'copra', 'have', 'reduced', 'the', 'margin', 'between', 'prices', 'in', 'the', 'two', 'countries', '.', 'Indonesia', "'s", 'copra', 'production', 'is', 'forecast', 'at', '1.32', 'mln', 'tonnes', 'in', 'calendar', '1987', ',', 'up', 'from', '1.30', 'mln', 'tonnes', 'in', '1986', '.']
```

----> End of sentence token module <----

pos_tagger():

Examples:

- Example 1:

DATA -

AUSTN FEB ANNUAL M3 MONEY SUPPLY RISES 11.2 PCT
The annual growth of Australia's m3
money supply rose by 11.2 pct in the year ended February
compared with January's 10.7 pct, the Reserve Bank said.
This was down from 14.0 pct in February last year.
In February m3 rose by 0.6 pct compared with 0.8 in January
and a February 1986 rise of 0.1 pct
More

OUTPUT -

----->Tagged Words<-----

```
[('AUSTN', 'NNP'), ('FEB', 'NNP'), ('ANNUAL', 'NNP'), ('M3', 'NNP'), ('MONEY', 'NNP'), ('SUPPLY', 'NNP'), ('RISES', 'NNP'), ('11.2', 'CD'), ('PCT', 'NNP'), ('The', 'DT'), ('annual', 'JJ'), ('growth', 'NN'), ('of', 'IN'), ('Australia', 'NNP'), ('s', 'POS'), ('m3', 'NN'), ('money', 'NN'), ('supply', 'NN'), ('rose', 'VBD'), ('by', 'IN'), ('11.2', 'CD'), ('pct', 'NN'), ('in', 'IN'), ('the', 'DT'), ('year', 'NN'), ('ended', 'VBD'), ('February', 'NNP'), ('compared', 'VBN'), ('with', 'IN'), ('January', 'NNP'), ('s', 'POS'), ('10.7', 'CD'), ('pct', 'NN'), (',', ','), ('the', 'DT'), ('Reserve', 'NNP'), ('Bank', 'NNP'), ('said', 'VBD'), ('.', '.'), ('This', 'DT'), ('was', 'VBD'), ('down', 'RB'), ('from', 'IN'), ('14.0', 'CD'), ('pct', 'NN'), ('in', 'IN'), ('February', 'NNP'), ('last', 'JJ'), ('year', 'NN'), ('.', '.'), ('In', 'IN'), ('February', 'NNP'), ('m3', 'NN'), ('rose', 'VBD'), ('by', 'IN'), ('0.6', 'CD'), ('pct', 'NNS'), ('compared', 'VBN'), ('with', 'IN'), ('0.8', 'CD'), ('in', 'IN'), ('January', 'NNP'), ('and', 'CC'), ('a', 'DT'), ('February', 'NNP'), ('1986', 'CD'), ('rise', 'NN'), ('of', 'IN'), ('0.1', 'CD'), ('pct', 'NN'), ('.', '.'), ('More', 'JJR')]
```

-----> End of Pos Tagger module <-----

- Example 2:

DATA -

R.R. DONNELLY &DNY> TO SELL CABLE SYSTEM
R.R. Donnelly and Sons Co said it
reached an agreement with Adams-Russell Co Inc &AAR> to sell
Adams its Rockford, Ill., cable system for an undisclosed
price.
R.R. Donnelly said it expects the deal to close in about
120 days, subject to regulatory approvals and other conditions.
The cable system serves about 51,000 subscribers in the
communities of Rockford, Loves Park, Machesney Park, Cherry
Valley, Morristown and Winnebago County, Ill., R.R. Donnelly
said.

OUTPUT -

----->Tagged Words<-----

```
[('R.R.', 'NNP'), ('.', '.'), ('DONNELLY', 'NNP'), ('&', 'CC'), ('lt', 'NN'),  
(';', ':'), ('DNY', 'NNP'), ('>', 'VBZ'), ('TO', 'NNP'), ('SELL', 'NNP'), ('  
'CABLE', 'NNP'), ('SYSTEM', 'NNP'), ('R.R.', 'NNP'), ('.', '.'), ('Donnelly',  
'NNP'), ('and', 'CC'), ('Sons', 'NNP'), ('Co', 'NNP'), ('said', 'VBD'), ('i  
t', 'PRP'), ('reached', 'VBD'), ('an', 'DT'), ('agreement', 'NN'), ('with',  
'IN'), ('Adams-Russell', 'NNP'), ('Co', 'NNP'), ('Inc', 'NNP'), ('&', 'CC'),  
'lt', 'NN'), (';', ':'), ('AAR', 'NNP'), ('>', 'VBZ'), ('to', 'TO'), ('sel  
l', 'VB'), ('Adams', 'NNP'), ('its', 'PRP$'), ('Rockford', 'NNP'), ('.', '.'), ('  
'Ill.', 'NNP'), ('.', '.'), ('cable', 'NN'), ('system', 'NN'), ('for', 'I  
N'), ('an', 'DT'), ('undisclosed', 'JJ'), ('price', 'NN'), ('.', '.'), ('R.  
R.', 'NNP'), ('.', '.'), ('Donnelly', 'NNP'), ('said', 'VBD'), ('it', 'PRP'),  
( 'expects', 'VBZ'), ('the', 'DT'), ('deal', 'NN'), ('to', 'TO'), ('close',  
'VB'), ('in', 'IN'), ('about', 'RB'), ('120', 'CD'), ('days', 'NNS'), ('.', '.'),  
''), ('subject', 'JJ'), ('to', 'TO'), ('regulatory', 'JJ'), ('approvals', 'I  
NNS'), ('and', 'CC'), ('other', 'JJ'), ('conditions', 'NNS'), ('.', '.'), ('  
The', 'DT'), ('cable', 'NN'), ('system', 'NN'), ('serves', 'VBZ'), ('about',  
'RB'), ('51,000', 'CD'), ('subscribers', 'NNS'), ('in', 'IN'), ('the', 'DT'  
'), ('communities', 'NNS'), ('of', 'IN'), ('Rockford', 'NNP'), ('.', '.'), ('  
Loves', 'NNP'), ('Park', 'NNP'), ('.', '.'), ('Machesney', 'NNP'), ('Park',  
'NNP'), ('.', '.'), ('Cherry', 'NNP'), ('Valley', 'NNP'), ('.', '.'), ('Morr  
istown', 'NNP'), ('and', 'CC'), ('Winnebago', 'NNP'), ('County', 'NNP'), ('.',  
'', '''), ('Ill.', 'NNP'), ('.', '.'), ('R.R.', 'NNP'), ('.', '.'), ('Donnelly  
'', 'NNP'), ('said', 'VBD'), ('.', '.')]
```

- Example 3:

DATA -

C.T.C. DEALER EXTENDS BID FOR CANADIAN TIRE

C.T.C. Dealer Holdings Ltd said it

extended its previously announced offer for 49 pct of <Canadian Tire Corp Ltd> to midnight on March 26.

CTC, a group of Canadian Tire dealers, which already owns 17 pct of Canadian Tire, is currently appealing in an Ontario court against a previously announced Ontario Securities Commission ruling blocking CTC's 272 mln dlr bid.

The blocked bid did not include non-voting shareholders, who hold 96 pct of Canadian Tire equity.

OUTPUT -

----->Tagged Words<-----

```
[('C.T.C.', 'NNP'), ('.', '.'), ('DEALER', 'NNP'), ('EXTENDS', 'NNP'), ('BID'  
'', 'NNP'), ('FOR', 'NNP'), ('CANADIAN', 'NNP'), ('TIRE', 'NNP'), ('C.T.C.', 'N  
NP'), ('.', '.'), ('Dealer', 'NNP'), ('Holdings', 'NNPS'), ('Ltd', 'NNP'), ('  
'said', 'VBD'), ('it', 'PRP'), ('extended', 'VBD'), ('its', 'PRP$'), ('previ  
ously', 'RB'), ('announced', 'VBN'), ('offer', 'NN'), ('for', 'IN'), ('49',  
'CD'), ('pct', 'NN'), ('of', 'IN'), ('&', 'CC'), ('lt', 'NN'), (';', ':'), ('  
'Canadian', 'NNP'), ('Tire', 'NNP'), ('Corp', 'NNP'), ('Ltd', 'NNP'), ('>',  
'NNP'), ('to', 'TO'), ('midnight', 'VB'), ('on', 'IN'), ('March', 'NNP'), ('  
26', 'CD'), ('of', 'IN'), ('CTC', 'NNP'), ('.', '.'), ('a', 'DT'), ('group', 'I  
NN'), ('of', 'IN'), ('Canadian', 'NNP'), ('Tire', 'NNP'), ('dealers', 'NNS')  
'', ''), ('which', 'WDT'), ('already', 'RB'), ('owns', 'VBZ'), ('17', 'I  
CD'), ('pct', 'NN'), ('of', 'IN'), ('Canadian', 'JJ'), ('Tire', 'NNP'), ('.',  
'', ''), ('is', 'VBZ'), ('currently', 'RB'), ('appealing', 'VBG'), ('in', 'IN'  
''), ('an', 'DT'), ('Ontario', 'NNP'), ('court', 'NN'), ('against', 'IN'), ('  
a', 'DT'), ('previously', 'RB'), ('announced', 'VBN'), ('Ontario', 'NNP'), ('  
'Securities', 'NNPS'), ('Commission', 'NNP'), ('ruling', 'VBG'), ('blocking'  
'', 'VBG'), ('CTC', 'NNP'), ('s', 'POS'), ('272', 'CD'), ('mln', 'JJ'), ('dlr'  
'', 'NN'), ('bid', 'NN'), ('.', '.'), ('The', 'DT'), ('blocked', 'JJ'), ('bid  
'', 'NN'), ('did', 'VBD'), ('not', 'RB'), ('include', 'VB'), ('non-voting', 'I  
JJ'), ('shareholders', 'NNS'), ('.', '.'), ('who', 'WP'), ('hold', 'VBP'), ('  
'96', 'CD'), ('pct', 'NN'), ('of', 'IN'), ('Canadian', 'JJ'), ('Tire', 'NNP'  
''), ('equity', 'NN'), ('.', '.')]
```

----- End of Pos Tagger module <-----

- Example 4:

DATA -

N.Z. UNEMPLOYMENT RATE 3.9 PCT IN DECEMBER QUARTER

New Zealand's unemployment rate was 3.9 pct of the workforce in the quarter ended December 31, unchanged from a revised 3.9 pct (preliminary 3.8 pct) in the previous quarter but slightly above 3.8 pct in the year-earlier quarter, the Statistics Department said.

The Department, citing the Household Labour Force Survey, said in a statement the number of unemployed in October-December 1986 was 61,500 against 60,500 in the September quarter and 60,500 a year earlier.

OUTPUT -

----->Tagged Words<-----

```
[('N.Z', 'NNP'), (',', ','), ('UNEMPLOYMENT', 'NNP'), ('RATE', 'NN'), ('3.9', 'CD'), ('PCT', 'NNP'), ('IN', 'NNP'), ('DECEMBER', 'NNP'), ('QUARTER', 'NNP'), ('New', 'NNP'), ('Zealand', 'NNP'), ('s', 'POS'), ('unemployment', 'NN'), ('rate', 'NN'), ('was', 'VBD'), ('3.9', 'CD'), ('pct', 'NN'), ('of', 'IN'), ('the', 'DT'), ('workforce', 'NN'), ('in', 'IN'), ('the', 'DT'), ('quarter', 'NN'), ('ended', 'VBD'), ('December', 'NNP'), ('31', 'CD'), ('', ''), ('unchanged', 'JJ'), ('from', 'IN'), ('a', 'DT'), ('revised', 'VBN'), ('3.9', 'CD'), ('pct', 'NN'), ('(', '('), ('preliminary', 'JJ'), ('3.8', 'CD'), ('pct', 'NN'), ('', ''), ('in', 'IN'), ('the', 'DT'), ('previous', 'JJ'), ('quarter', 'NN'), ('but', 'CC'), ('slightly', 'RB'), ('above', 'IN'), ('3.8', 'CD'), ('pct', 'NN'), ('in', 'IN'), ('the', 'DT'), ('year-earlier', 'JJ'), ('quarter', 'NN'), ('the', 'DT'), ('Statistics', 'NNPS'), ('Department', 'NNP'), ('said', 'VBD'), ('The', 'DT'), ('Department', 'NNP'), ('citing', 'VBG'), ('the', 'DT'), ('Household', 'NNP'), ('Labour', 'NNP'), ('Force', 'NNP'), ('Survey', 'NNP'), ('said', 'VBD'), ('in', 'IN'), ('a', 'DT'), ('statement', 'NN'), ('the', 'DT'), ('number', 'NN'), ('of', 'IN'), ('unemployed', 'JJ'), ('in', 'IN'), ('October', 'NNP'), ('December', 'NNP'), ('1986', 'CD'), ('was', 'VBD'), ('61,500', 'CD'), ('against', 'IN'), ('60,500', 'CD'), ('in', 'IN'), ('the', 'DT'), ('September', 'NNP'), ('quarter', 'NN'), ('and', 'CC'), ('60,500', 'CD'), ('a', 'DT'), ('year', 'NN'), ('earlier', 'RBR'), ('', '')]
```

-----> End of Pos Tagger module <-----

- Example 5:

DATA -

M.A. HANNA CO <HNM> TO ACQUIRE <PMS CONSOLIDATED>

M.A. Hanna Co said it has reached definitive agreement to purchase PMS Consolidated, a privately owned manufacturer of colorants for the plastics industry.

M.A. Hanna said the purchase, which is subject to certain conditions, is scheduled to close around the end of October. Hanna announced last August that it was in negotiations to purchase the company.

Headquartered in Somerset, N.J., PMS reported over 70 mln dls in revenues in its last fiscal year, which ended May 1987.

OUTPUT -

----->Tagged Words<-----

```
[('M.A.', 'NNP'), (',', ','), ('HANNA', 'NNP'), ('CO', 'NNP'), ('&', 'CC'), ('lt', 'NN'), (';', ':'), ('HNM', 'NNP'), ('>', 'NNP'), ('TO', 'NNP'), ('ACQURE', 'NNP'), ('&', 'CC'), ('lt', 'NN'), (';', ':'), ('PMS', 'NNP'), ('CONSOLIDATED', 'NNP'), ('>', 'NNP'), ('M.A.', 'NNP'), (',', ','), ('Hanna', 'NNP'), ('Co', 'NNP'), ('said', 'VBD'), ('it', 'PRP'), ('has', 'VBZ'), ('reached', 'VBN'), ('definitive', 'JJ'), ('agreement', 'NN'), ('to', 'TO'), ('purchase', 'VB'), ('PMS', 'NNP'), ('Consolidated', 'NNP'), (',', ','), ('a', 'DT'), ('privately', 'RB'), ('owned', 'VBN'), ('manufacturer', 'NN'), ('of', 'IN'), ('colorants', 'NNS'), ('for', 'IN'), ('the', 'DT'), ('plastics', 'NNS'), ('industry', 'NN'), (',', ','), ('M.A.', 'NNP'), (',', ','), ('Hanna', 'NNP'), ('said', 'VBD'), ('the', 'DT'), ('purchase', 'NN'), (',', ','), ('which', 'WD T'), ('is', 'VBZ'), ('subject', 'JJ'), ('to', 'TO'), ('certain', 'JJ'), ('conditions', 'NNS'), (',', ','), ('is', 'VBZ'), ('scheduled', 'VBN'), ('to', 'TO'), ('close', 'VB'), ('around', 'IN'), ('the', 'DT'), ('end', 'NN'), ('of', 'IN'), ('October', 'NNP'), (',', ','), ('Hanna', 'NNP'), ('announced', 'VBD'), ('last', 'JJ'), ('August', 'NNP'), ('that', 'IN'), ('it', 'PRP'), ('was', 'VBD'), ('in', 'IN'), ('negotiations', 'NNS'), ('to', 'TO'), ('purchase', 'VB'), ('the', 'DT'), ('company', 'NN'), (',', ','), ('Headquartered', 'VBN'), ('in', 'IN'), ('Somerset', 'NNP'), (',', ','), ('N.J.', 'NNP'), (',', ','), ('PMS', 'NNP'), ('reported', 'VBD'), ('over', 'IN'), ('70', 'CD'), ('mln', 'NNS'), ('dls', 'NN'), ('in', 'IN'), ('revenues', 'NNS'), ('in', 'IN'), ('its', 'PRP$'), ('last', 'JJ'), ('fiscal', 'JJ'), ('year', 'NN'), (',', ','), ('which', 'WDT'), ('ended', 'VBD'), ('May', 'NNP'), ('1987', 'CD'), (',', ','), ('.')] ]
```

-----> End of Pos Tagger module <-----

- Example 6:

DATA-

INDONESIA UNLIKELY TO IMPORT PHILIPPINES COPRA

Indonesia is unlikely to import copra from the Philippines in 1987 after importing 30,000 tonnes in 1986, the U.S. Embassy's annual agriculture report said.

The report said the 31 pct devaluation of the Indonesian rupiah, an increase in import duties on copra and increases in the price of Philippines copra have reduced the margin between prices in the two countries.

Indonesia's copra production is forecast at 1.32 mln tonnes in calendar 1987, up from 1.30 mln tonnes in 1986.

OUTPUT -

----->Tagged Words<-----

```
[('INDONESIA', 'NNP'), ('UNLIKELY', 'NNP'), ('TO', 'NNP'), ('IMPORT', 'NNP'), ('PHILIPPINES', 'NNP'), ('COPRA', 'NNP'), ('Indonesia', 'NNP'), ('is', 'VBZ'), ('unlikely', 'JJ'), ('to', 'TO'), ('import', 'VB'), ('copra', 'NN'), ('from', 'IN'), ('the', 'DT'), ('Philippines', 'NNPS'), ('in', 'IN'), ('1987', 'CD'), ('after', 'IN'), ('importing', 'VBG'), ('30,000', 'CD'), ('tonnes', 'NNS'), ('in', 'IN'), ('1986', 'CD'), (',', ','), ('the', 'DT'), ('U.S.', 'NNP'), ('Embassy', 'NNP'), ('s', 'POS'), ('annual', 'JJ'), ('agriculture', 'NN'), ('report', 'NN'), ('said', 'VBD'), (',', ','), ('The', 'DT'), ('report', 'NN'), ('said', 'VBD'), ('the', 'DT'), ('31', 'CD'), ('pct', 'JJ'), ('devaluation', 'NN'), ('of', 'IN'), ('the', 'DT'), ('Indonesian', 'NNP'), ('rupiah', 'NN'), (',', ','), ('an', 'DT'), ('increase', 'NN'), ('in', 'IN'), ('import', 'JJ'), ('duties', 'NNS'), ('on', 'IN'), ('copra', 'NN'), ('and', 'CC'), ('increases', 'NNS'), ('in', 'IN'), ('the', 'DT'), ('price', 'NN'), ('of', 'IN'), ('Philippines', 'NNPS'), ('copra', 'NNS'), ('have', 'VBP'), ('reduced', 'VBN'), ('the', 'DT'), ('margin', 'NN'), ('between', 'IN'), ('prices', 'NNS'), ('in', 'IN'), ('the', 'DT'), ('two', 'CD'), ('countries', 'NNS'), (',', ','), ('Indonesia', 'NNP'), ('s', 'POS'), ('copra', 'NN'), ('production', 'NN'), ('is', 'VBZ'), ('forecast', 'VBN'), ('at', 'IN'), ('1.32', 'CD'), ('mln', 'NN'), ('tonnes', 'NNS'), ('in', 'IN'), ('calendar', 'NN'), ('1987', 'CD'), (',', ','), ('up', 'RB'), ('from', 'IN'), ('1.30', 'CD'), ('mln', 'NN'), ('tonnes', 'NNS'), ('in', 'IN'), ('1986', 'CD'), (',', ','), ('.')] ]
```

-----> End of Pos Tagger module <-----

measuredEntityDetection:

Examples:

- Example 1:

DATA -

AUSTN FEB ANNUAL M3 MONEY SUPPLY RISES 11.2 PCT

The annual growth of Australia's m3
money supply rose by 11.2 pct in the year ended February
compared with January's 10.7 pct, the Reserve Bank said.
This was down from 14.0 pct in February last year.
In February m3 rose by 0.6 pct compared with 0.8 in January
and a February 1986 rise of 0.1 pct
More

OUTPUT -

---->Measured Entity detected Word list<----

```
('AUSTN', 'NNP')
('FEB', 'NNP')
('ANNUAL', 'NNP')
('M3', 'NNP')
('MONEY', 'NNP')
('SUPPLY', 'NNP')
('RISES', 'NNP')
('ME 11.2/CD PCT/NNP')
('The', 'DT')
('annual', 'JJ')
('growth', 'NN')
('of', 'IN')
('Australia', 'NNP')
('s', 'POS')
('m3', 'NN')
('money', 'NN')
('supply', 'NN')
('rose', 'VBD')
('by', 'IN')
('ME 11.2/CD pct/NN')
('in', 'IN')
('the', 'DT')
('year', 'NN')
('ended', 'VBD')
('February', 'NNP')
('compared', 'VBN')
('with', 'IN')
('January', 'NNP')
('s', 'POS')
('ME 10.7/CD pct/NN')
(' , , ,')
```

- Example 2:

DATA -

R.R. DONNELLY &DNY> TO SELL CABLE SYSTEM

R.R. Donnelly and Sons Co said it
reached an agreement with Adams-Russell Co Inc &AAR> to sell
Adams its Rockford, Ill., cable system for an undisclosed
price.

R.R. Donnelly said it expects the deal to close in about
120 days, subject to regulatory approvals and other conditions.

The cable system serves about 51,000 subscribers in the
communities of Rockford, Loves Park, Machesney Park, Cherry
Valley, Morristown and Winnebago County, Ill., R.R. Donnelly
Said.

OUTPUT -

```
('the', 'DT')
('deal', 'NN')
('to', 'TO')
('close', 'VB')
('in', 'IN')
('about', 'RB')
(ME 120/CD days/NNS)
(', ', ', ')
('subject', 'JJ')
('to', 'TO')
('regulatory', 'JJ')
('approvals', 'NNS')
('and', 'CC')
('other', 'JJ')
('conditions', 'NNS')
('.', '. ')
('The', 'DT')
('cable', 'NN')
('system', 'NN')
('serves', 'VBZ')
('about', 'RB')
(ME 51,000/CD subscribers/NNS)
('in', 'IN')
('the', 'DT')
('communities', 'NNS')
```

- Example 3:

DATA -

C.T.C. DEALER EXTENDS BID FOR CANADIAN TIRE

C.T.C. Dealer Holdings Ltd said it extended its previously announced offer for 49 pct of <Canadian Tire Corp Ltd> to midnight on March 26.

CTC, a group of Canadian Tire dealers, which already owns 17 pct of Canadian Tire, is currently appealing in an Ontario court against a previously announced Ontario Securities Commission ruling blocking CTC's 272 mln dlr bid.

The blocked bid did not include non-voting shareholders, who hold 96 pct of Canadian Tire equity.

OUTPUT -

```
(ME 17/CD pct/NN)
('of', 'IN')
('Canadian', 'JJ')
('Tire', 'NNP')
(',', ',')
('is', 'VBZ')
('currently', 'RB')
('appealing', 'VBG')
('in', 'IN')
('an', 'DT')
('Ontario', 'NNP')
('court', 'NN')
('against', 'IN')
('a', 'DT')
('previously', 'RB')
('announced', 'VBN')
('Ontario', 'NNP')
('Securities', 'NNPS')
('Commission', 'NNP')
('ruling', 'VBG')
('blocking', 'VBG')
('CTC', 'NNP')
('s', 'POS')
(ME 272/CD mln/JJ dlr/NN bid/NN)
('.', '.')
('The', 'DT')
('blocked', 'JJ')
('bid', 'NN')
('did', 'VBD')
('not', 'RB')
('include', 'VB')
('non-voting', 'JJ')
('shareholders', 'NNS')
(',', ',')
('who', 'WP')
('hold', 'VBP')
(ME 96/CD pct/NN)
```

- Example 4:

DATA -

N.Z. UNEMPLOYMENT RATE 3.9 PCT IN DECEMBER QUARTER

New Zealand's unemployment rate was

3.9 pct of the workforce in the quarter ended December 31, unchanged from a revised 3.9 pct (preliminary 3.8 pct) in the previous quarter but slightly above 3.8 pct in the year-earlier quarter, the Statistics Department said.

The Department, citing the Household Labour Force Survey, said in a statement the number of unemployed in October-December 1986 was 61,500 against 60,500 in the September quarter and 60,500 a year earlier.

OUTPUT -

---->Mesured Enitity detected Word list<----

```
('N.Z', 'NNP')
('.', '.')
('UNEMPLOYMENT', 'NNP')
('RATE', 'NN')
('ME 3.9/CD PCT/NNP')
('IN', 'NNP')
('DECEMBER', 'NNP')
('QUARTER', 'NNP')
('New', 'NNP')
('Zealand', 'NNP')
(''s', 'POS')
('unemployment', 'NN')
('rate', 'NN')
('was', 'VBD')
('ME 3.9/CD pct/NN')
('of', 'IN')
('the', 'DT')
('workforce', 'NN')
('in', 'IN')
('the', 'DT')
('quarter', 'NN')
('ended', 'VBD')
('December', 'NNP')
('31', 'CD')
(',', ',')
('unchanged', 'JJ')
('from', 'IN')
('a', 'DT')
('revised', 'VBN')
('ME 3.9/CD pct/NN')
('(', '(')
('preliminary', 'JJ')
('ME 3.8/CD pct/NN')
```

- Example 5:

DATA -

M.A. HANNA CO <HNM> TO ACQUIRE <PMS CONSOLIDATED>

M.A. Hanna Co said it has reached definitive agreement to purchase PMS Consolidated, a privately owned manufacturer of colorants for the plastics industry.

M.A. Hanna said the purchase, which is subject to certain conditions, is scheduled to close around the end of October. Hanna announced last August that it was in negotiatitons to purchase the company.

Headquartered in Somerset, N.J., PMS reported over 70 mln dlrs in revenues in its last fiscal year, which ended May 1987.

OUTPUT -


```
( 'last', 'JJ' )
( 'August', 'NNP' )
( 'that', 'IN' )
( 'it', 'PRP' )
( 'was', 'VBD' )
( 'in', 'IN' )
( 'negotiations', 'NNS' )
( 'to', 'TO' )
( 'purchase', 'VB' )
( 'the', 'DT' )
( 'company', 'NN' )
( '.', '.' )
( 'Headquartered', 'VBN' )
( 'in', 'IN' )
( 'Somerset', 'NNP' )
( ',', ',' )
( 'N.J.', 'NNP' )
( ',', ',' )
( 'PMS', 'NNP' )
( 'reported', 'VBD' )
( 'over', 'IN' )
( ME 70/CD mln/NNS )
( 'dlrs', 'NN' )
( 'in', 'IN' )
( 'revenues', 'NNS' )
( 'in', 'IN' )
( 'its', 'PRP$' )
( 'last', 'JJ' )
```

- Example 6:

DATA-

INDONESIA UNLIKELY TO IMPORT PHILIPPINES COPRA

Indonesia is unlikely to import copra from the Philippines in 1987 after importing 30,000 tonnes in 1986, the U.S. Embassy's annual agriculture report said.

The report said the 31 pct devaluation of the Indonesian rupiah, an increase in import duties on copra and increases in the price of Philippines copra have reduced the margin between prices in the two countries.

Indonesia's copra production is forecast at 1.32 mln tonnes in calendar 1987, up from 1.30 mln tonnes in 1986.

OUTPUT -

```
( 'in', 'IN' )
( 'the', 'DT' )
( ME two/CD countries/NNS )
( '.', '.' )
( 'Indonesia', 'NNP' )
( 's', 'POS' )
( 'copra', 'NN' )
( 'production', 'NN' )
( 'is', 'VBZ' )
( 'forecast', 'VBN' )
( 'at', 'IN' )
( ME 1.32/CD mln/NN )
( 'tonnes', 'NNS' )
( 'in', 'IN' )
( 'calendar', 'NN' )
( '1987', 'CD' )
( ',', ',' )
( 'up', 'RB' )
( 'from', 'IN' )
( ME 1.30/CD mln/NN )
( 'tonnes', 'NNS' )
( 'Philippines', 'NNPS' )
( 'in', 'IN' )
( '1987', 'CD' )
( 'after', 'IN' )
( 'importing', 'VBG' )
( ME 30,000/CD tonnes/NNS )
( 'in', 'IN' )
( '1986', 'CD' )
( ',', ',' )
( 'the', 'DT' )
( 'U.S.', 'NNP' )
( 'Embassy', 'NNP' )
( 's', 'POS' )
```

dateRecognizer():

Sample data in NLTK Reuters was not vast enough to showcase the capability of date Parser and date recognizer.

Below are several types of data given to the modules. First is the NLTK Reuters library data and the second is a paragraph written by myself with 8 date strings which I want daterecognizer() to recognize and parse.

Examples:

- Example 1:

DATA -

AUSTN FEB ANNUAL M3 MONEY SUPPLY RISES 11.2 PCT
The annual growth of Australia's m3
money supply rose by 11.2 pct in the year ended February
compared with January's 10.7 pct, the Reserve Bank said.
This was down from 14.0 pct in February last year.
In February m3 rose by 0.6 pct compared with 0.8 in January
and a February 1986 rise of 0.1 pct
More

OUTPUT - February 1986

```
----> Date Recogniser module list<----
```

February 1986

```
----> End of Date Recogniser Module <----
```

- Example 2:

DATA -

"Not so long ago, while I was in Montreal. That day the date was 9/30/2020. I was heading to a party. Earlier I thought it was on August 10 2020. But I was mistaken, later I found that my calendar was old, instead of showing the month of December 2019, it was showing the date of 31st December. I remember I came to Montreal on the 20th of August. Earlier I had planned to come on September 24th. I didn't come on Thirty first of the Second of October 2020, because I was unable to get tickets on that day. This is getting boring and today is the 21st of May of year 1053."

OUTPUT -

```
----> Date Recogniser module list<----
```

```
9/30/2020  
August 10 2020  
20th of August  
September 24th  
Second of October 2020  
21st of May 1053
```

```
----> End of Date Recogniser Module <----
```

- Example 3:

DATA -

R.R. DONNELLY &DNY> TO SELL CABLE SYSTEM

R.R. Donnelly and Sons Co said it reached an agreement with Adams-Russell Co Inc &AAR> to sell Adams its Rockford, Ill., cable system for an undisclosed price.

R.R. Donnelly said it expects the deal to close in about 120 days, subject to regulatory approvals and other conditions.

The cable system serves about 51,000 subscribers in the communities of Rockford, Loves Park, Machesney Park, Cherry Valley, Morristown and Winnebago County, Ill., R.R. Donnelly said.

OUTPUT -

```
----> Date Recogniser module list<----
```

```
----> End of Date Recogniser Module <----
```

- Example 4:
DATA -

C.T.C. DEALER EXTENDS BID FOR CANADIAN TIRE

C.T.C. Dealer Holdings Ltd said it extended its previously announced offer for 49 pct of &Canadian Tire Corp Ltd> to midnight on March 26.

CTC, a group of Canadian Tire dealers, which already owns 17 pct of Canadian Tire, is currently appealing in an Ontario court against a previously announced Ontario Securities Commission ruling blocking CTC's 272 mln dlr bid.

The blocked bid did not include non-voting shareholders, who hold 96 pct of Canadian Tire equity.

OUTPUT -

----> Date Recogniser module list<----

March 26

----> End of Date Recogniser Module <----

- Example 5:

DATA -

N.Z. UNEMPLOYMENT RATE 3.9 PCT IN DECEMBER QUARTER

New Zealand's unemployment rate was 3.9 pct of the workforce in the quarter ended December 31, unchanged from a revised 3.9 pct (preliminary 3.8 pct) in the previous quarter but slightly above 3.8 pct in the year-earlier quarter, the Statistics Department said.

The Department, citing the Household Labour Force Survey, said in a statement the number of unemployed in October-December 1986 was 61,500 against 60,500 in the September quarter and 60,500 a year earlier.

OUTPUT -

----> Date Recogniser module list<----

December 31

December 1986

----> End of Date Recogniser Module <----

- Example 6:

DATA -

M.A. HANNA CO <HNM> TO ACQUIRE <PMS CONSOLIDATED>

M.A. Hanna Co said it has reached definitive agreement to purchase PMS Consolidated, a privately owned manufacturer of colorants for the plastics industry.

M.A. Hanna said the purchase, which is subject to certain conditions, is scheduled to close around the end of October. Hanna announced last August that it was in negotiations to purchase the company.

Headquartered in Somerset, N.J., PMS reported over 70 mln dlrs in revenues in its last fiscal year, which ended May 1987.

OUTPUT -

----> Date Recogniser module list<----

May 1987

----> End of Date Recogniser Module <----

- Example 7:

DATA-

INDONESIA UNLIKELY TO IMPORT PHILIPPINES COPRA

Indonesia is unlikely to import copra from the Philippines in 1987 after importing 30,000 tonnes in 1986, the U.S. Embassy's annual agriculture report said.

The report said the 31 pct devaluation of the Indonesian rupiah, an increase in import duties on copra and increases in the price of Philippines copra have reduced the margin between prices in the two countries.

Indonesia's copra production is forecast at 1.32 mln tonnes in calendar 1987, up from 1.30 mln tonnes in 1986.

OUTPUT -

----> Date Recogniser module list<----

1986

1986

----> End of Date Recogniser Module <----

dateParserCFG():

Examples:

Example 1:

DATA -

AUSTN FEB ANNUAL M3 MONEY SUPPLY RISES 11.2 PCT

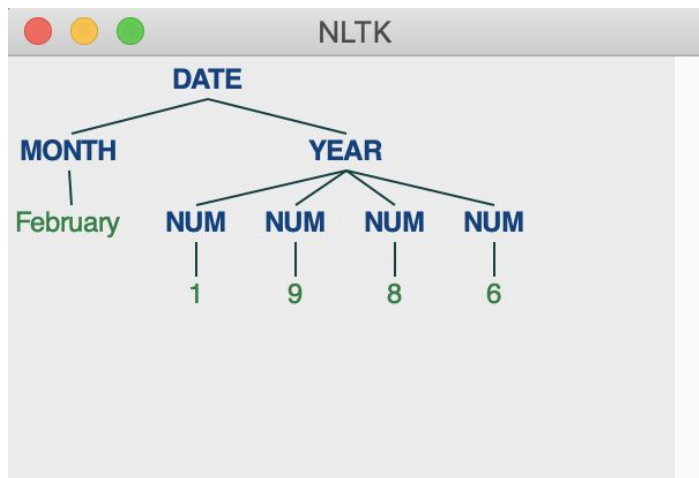
The annual growth of Australia's m3 money supply rose by 11.2 pct in the year ended February compared with January's 10.7 pct, the Reserve Bank said.

This was down from 14.0 pct in February last year.

In February m3 rose by 0.6 pct compared with 0.8 in January and a February 1986 rise of 0.1 pct

More

OUTPUT - February 1986

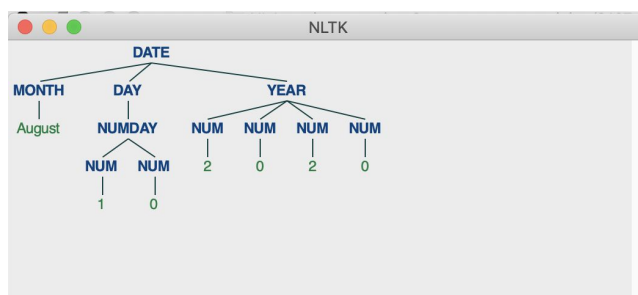
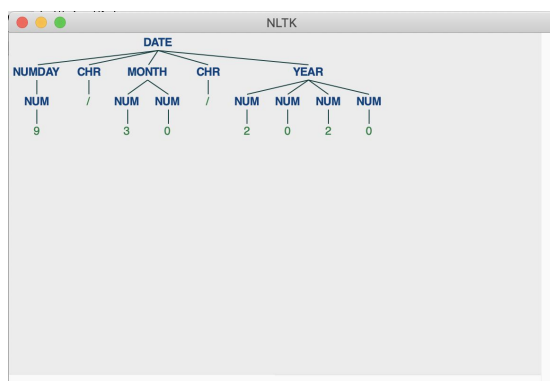


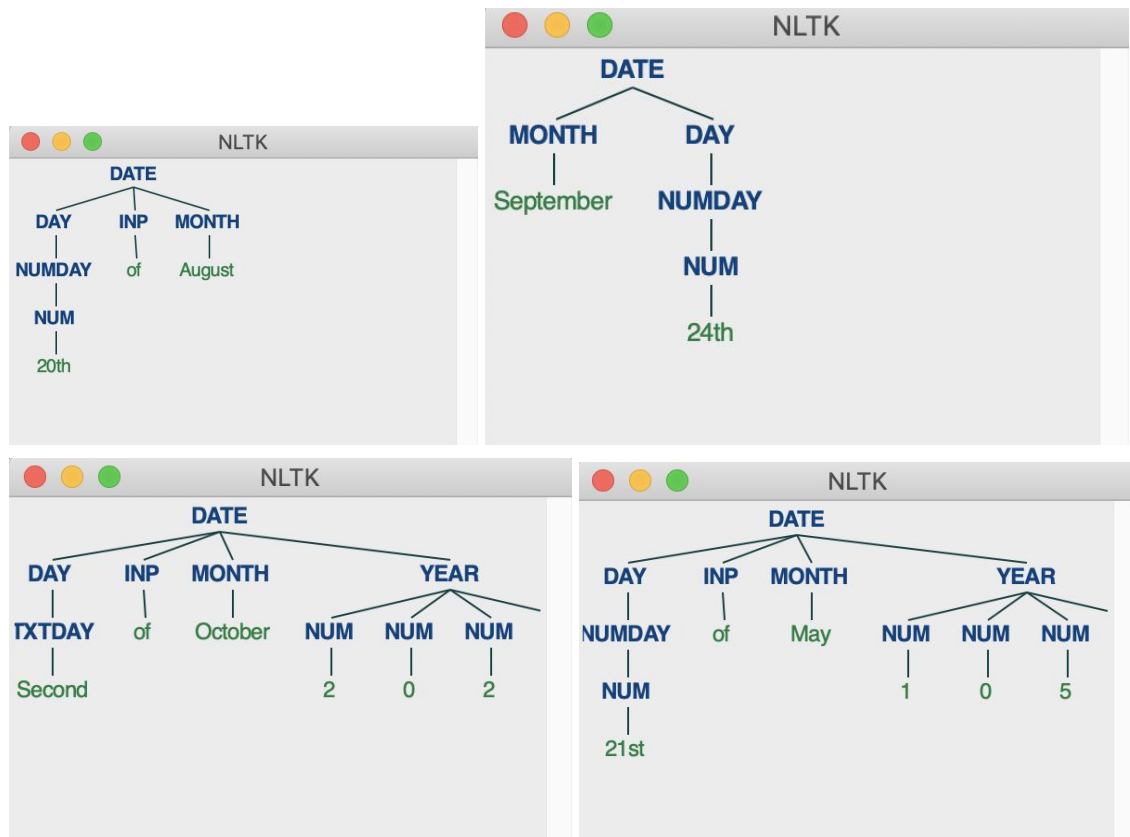
Example 2:

DATA -

"Not so long ago, while I was in Montreal. That day the date was 9/30/2020. I was heading to a party. Earlier I thought it was on August 10 2020. But I was mistaken, later I found that my calendar was old, instead of showing the month of December 2019, it was showing the date of 31st December. I remember I came to Montreal on the 20th of August. Earlier I had planned to come on September 24th. I didn't come on Thirty first of the Second of October 2020, because I was unable to get tickets on that day. This is getting boring and today is the 21st of May of year 1053."

OUTPUT -





- Example 3:

DATA -

R.R. DONNELLY <DNY> TO SELL CABLE SYSTEM
R.R. Donnelly and Sons Co said it
reached an agreement with Adams-Russell Co Inc <AAR> to sell
Adams its Rockford, Ill., cable system for an undisclosed
price.
R.R. Donnelly said it expects the deal to close in about
120 days, subject to regulatory approvals and other conditions.
The cable system serves about 51,000 subscribers in the
communities of Rockford, Loves Park, Machesney Park, Cherry
Valley, Morristown and Winnebago County, Ill., R.R. Donnelly
said.

OUTPUT -

---->Parsed Dates<----

----> End of Date Parser module <----

- Example 4:

DATA -

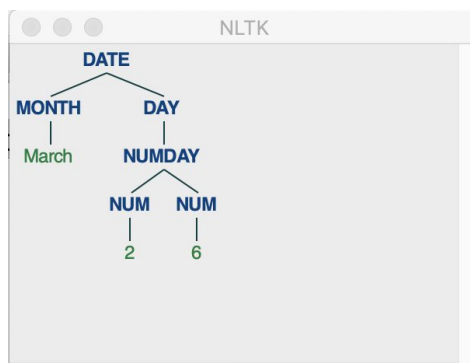
C.T.C. DEALER EXTENDS BID FOR CANADIAN TIRE

C.T.C. Dealer Holdings Ltd said it extended its previously announced offer for 49 pct of <Canadian Tire Corp Ltd> to midnight on March 26.

CTC, a group of Canadian Tire dealers, which already owns 17 pct of Canadian Tire, is currently appealing in an Ontario court against a previously announced Ontario Securities Commission ruling blocking CTC's 272 mln dlr bid.

The blocked bid did not include non-voting shareholders, who hold 96 pct of Canadian Tire equity.

OUTPUT -



----->Parsed Dates<-----

(DATE (MONTH March) (DAY (NUMDAY (NUM 2) (NUM 6))))

-----> End of Date Parser module <-----

END

- Example 5:

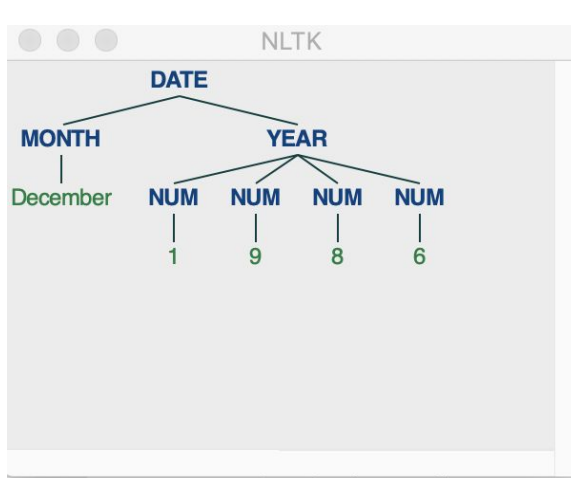
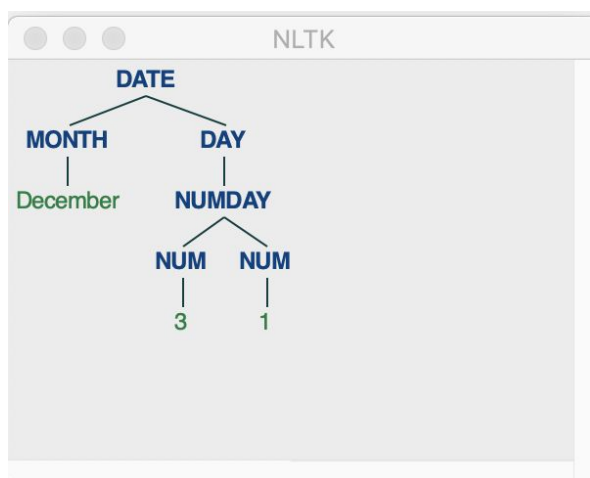
DATA -

N.Z. UNEMPLOYMENT RATE 3.9 PCT IN DECEMBER QUARTER

New Zealand's unemployment rate was 3.9 pct of the workforce in the quarter ended December 31, unchanged from a revised 3.9 pct (preliminary 3.8 pct) in the previous quarter but slightly above 3.8 pct in the year-earlier quarter, the Statistics Department said.

The Department, citing the Household Labour Force Survey, said in a statement the number of unemployed in October-December 1986 was 61,500 against 60,500 in the September quarter and 60,500 a year earlier.

OUTPUT -



----->Parsed Dates<-----

```
(DATE (MONTH December) (DAY (NUMDAY (NUM 3) (NUM 1))))  
(DATE (MONTH December) (YEAR (NUM 1) (NUM 9) (NUM 8) (NUM 6)))
```

-----> End of Date Parser module <-----

- Example 6:

DATA -

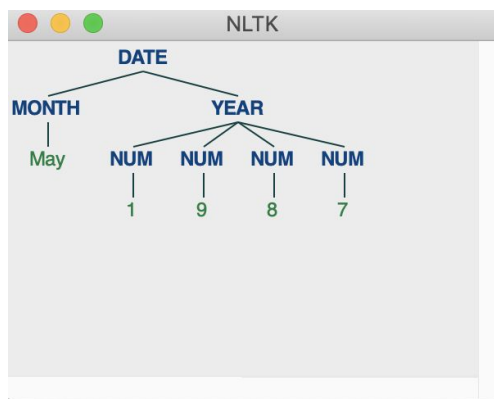
M.A. HANNA CO <HNM> TO ACQUIRE <PMS CONSOLIDATED>

M.A. Hanna Co said it has reached definitive agreement to purchase PMS Consolidated, a privately owned manufacturer of colorants for the plastics industry.

M.A. Hanna said the purchase, which is subject to certain conditions, is scheduled to close around the end of October. Hanna announced last August that it was in negotiations to purchase the company.

Headquartered in Somerset, N.J., PMS reported over 70 mln dlrs in revenues in its last fiscal year, which ended May 1987.

OUTPUT -



----->Parsed Dates<-----

```
(DATE (MONTH May) (YEAR (NUM 1) (NUM 9) (NUM 8) (NUM 7)))
```

-----> End of Date Parser module <-----

-----FND-----

- Example 7:

DATA-

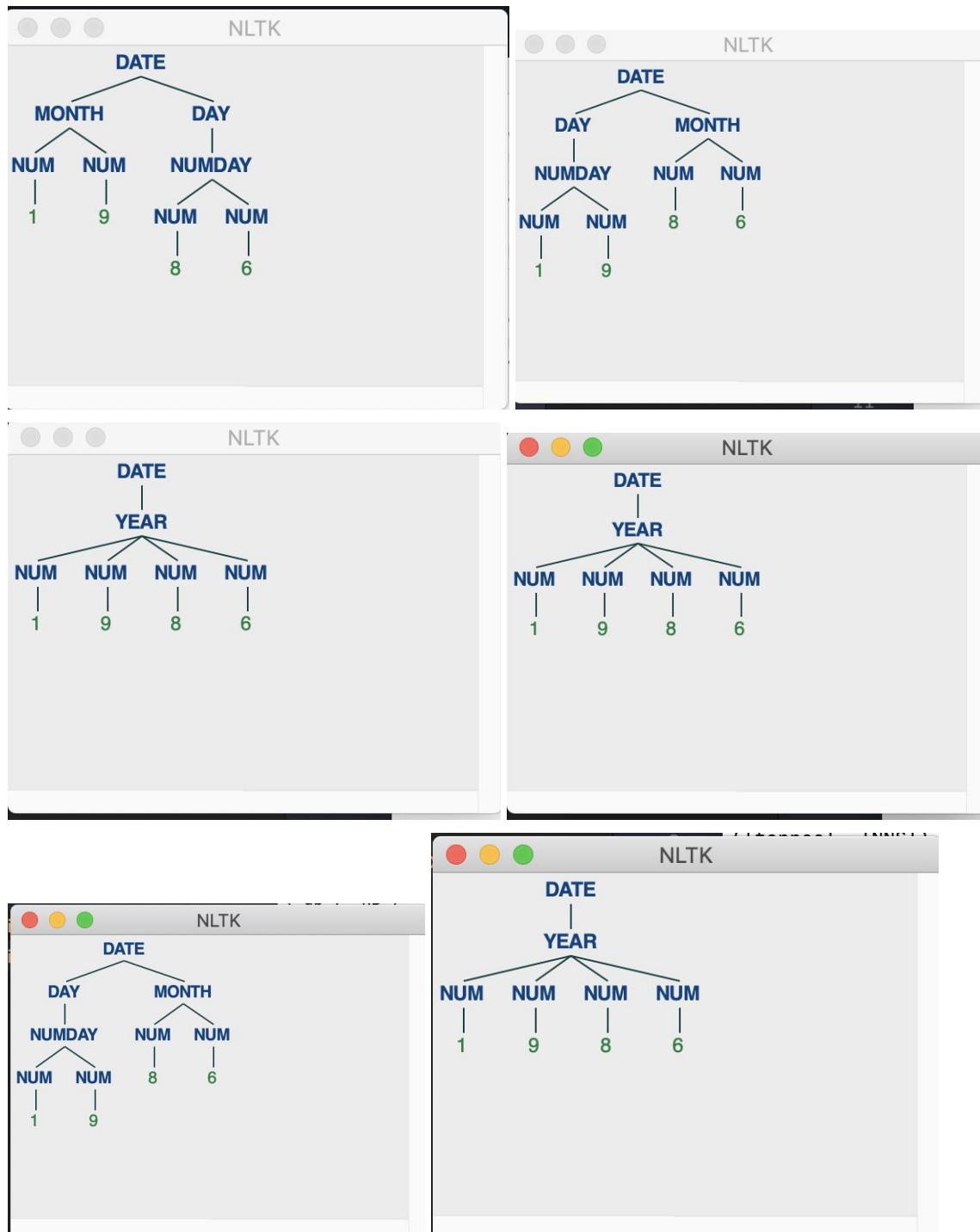
INDONESIA UNLIKELY TO IMPORT PHILIPPINES COPRA
Indonesia is unlikely to import copra

from the Philippines in 1987 after importing 30,000 tonnes in 1986, the U.S. Embassy's annual agriculture report said.

The report said the 31 pct devaluation of the Indonesian rupiah, an increase in import duties on copra and increases in the price of Philippines copra have reduced the margin between prices in the two countries.

Indonesia's copra production is forecast at 1.32 mln tonnes in calendar 1987, up from 1.30 mln tonnes in 1986.

OUTPUT -

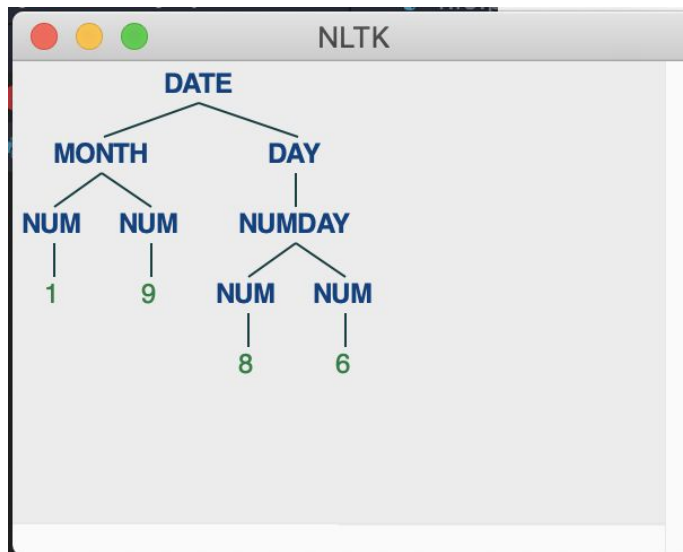


----->Parsed Dates<-----

```
(DATE (MONTH (NUM 1) (NUM 9)) (DAY (NUMDAY (NUM 8) (NUM 6))))  
(DATE (DAY (NUMDAY (NUM 1) (NUM 9))) (MONTH (NUM 8) (NUM 6)))  
(DATE (YEAR (NUM 1) (NUM 9) (NUM 8) (NUM 6)))  
(DATE (MONTH (NUM 1) (NUM 9)) (DAY (NUMDAY (NUM 8) (NUM 6))))  
(DATE (DAY (NUMDAY (NUM 1) (NUM 9))) (MONTH (NUM 8) (NUM 6)))  
(DATE (YEAR (NUM 1) (NUM 9) (NUM 8) (NUM 6)))
```

-----> End of Date Parser module <-----

END



named_entity():

Examples:

Example 1:

DATA -
OUTPUT -

Example 2:

DATA -
OUTPUT -

Example 3:

DATA -
OUTPUT -

Example 4:

DATA -
OUTPUT -

Example 5:

DATA -
OUTPUT -

END