

COMP 6751 Natural Language Analysis

Project 1 Report 2 (Demo)

Student- Amrita Awasthi (40203422)

Table of Contents-

1. Input and Outputs.....	2
1.1. Test Case 1: “reuters/training/267”.....	2
1.2. Test Case 2: “reuters/training/279”.....	3
1.3. Test Case 3: “reuters/training/9865”.....	4
1.4. Test Case 4: “reuters/training/9880”.....	5
2. Interesting Case.....	6
2.1. Test Case 5: “reuters/training/6”.....	6

1. Input and Outputs

The following test cases shows the input and outputs for the modules tokenization, sentence splitting, POS (Part of Speech) tagging, gazetteer annotation, named entity recognition (NER), and measured entity recognition.

1.1 Test Case 1

Input 1- Reuter's text training/267

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 1-

Tokenization-

```
Tokens: ['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', '.']
```

Sentence Splitting-

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

POS Tagging-

```
POS Tags: [('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', 'CD'), (',', ','), ('000', 'CD'), ('tonnes', 'NNS'), ('in', 'IN'), ('1986', 'CD'), (',', ','), ('the', 'DT'), ('U', 'NNP'), ('.', '.'), ('S', 'NNP'), ('.', '.'), ('Embassy's', 'NNP'), ('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's', 'NNP'), ('copra', 'JJ'), ('production', 'NN'), ('is', 'VBZ'), ('forecast', 'VBN'), ('at', 'IN'), ('1', 'CD'), ('.', '.'), ('32', 'CD'), ('mln', 'JJ'), ('tonnes', 'NNS'), ('in', 'IN'), ('calendar', 'NN'), ('1987', 'CD'), (',', ','), ('up', 'RB'), ('from', 'IN'), ('1', 'CD'), ('.', '.'), ('30', 'CD'), ('mln', 'JJ'), ('tonnes', 'NNS'), ('in', 'IN'), ('1986', 'CD'), ('.', '.'))]
```

Gazetteer Annotation-

```
Annotated Tokens: [('INDONESIA', 'O'), ('UNLIKELY', 'O'), ('TO', 'O'), ('IMPORT', 'O'), ('PHILIPPINES', 'O'), ('COPRA', 'O'), ('Indonesia', 'country'), ('is', 'O'), ('unlikely', 'O'), ('to', 'O'), ('import', 'O'), ('copra', 'crop'), ('from', 'O'), ('the', 'O'), ('Philippines', 'country'), ('in', 'O'), ('1987', 'year'), ('after', 'O'), ('importing', 'O'), ('30', 'O'), (',', 'O'), ('000', 'O'), ('tonnes', 'unit'), ('in', 'O'), ('1986', 'year'), ('.', 'O'), ('the', 'O'), ('U', 'O'), ('.', 'O'), ('S', 'O'), ('.', 'O'), ('Embassy's', 'O'), ('annual', 'O'), ('agriculture', 'O'), ('report', 'O'), ('said', 'O'), ('.', 'O'), ('The', 'O'), ('report', 'O'), ('said', 'O'), ('the', 'O'), ('31', 'O'), ('pct', 'O'), ('devaluation', 'O'), ('of', 'O'), ('the', 'O'), ('Indonesian', 'O'), ('rupiah', 'O'), ('.', 'O'), ('an', 'O'), ('increase', 'O'), ('in', 'O'), ('import', 'O'), ('duties', 'O'), ('on', 'O'), ('copra', 'crop'), ('and', 'O'), ('increases', 'O'), ('in', 'O'), ('the', 'O'), ('price', 'O'), ('of', 'O'), ('Philippines', 'country'), ('copra', 'crop'), ('have', 'O'), ('reduced', 'O'), ('the', 'O'), ('margin', 'O'), ('between', 'O'), ('prices', 'O'), ('in', 'O'), ('the', 'O'), ('two', 'O'), ('countries', 'O'), ('.', 'O'), ('Indonesia's', 'O'), ('copra', 'crop'), ('production', 'O'), ('is', 'O'), ('forecast', 'O'), ('at', 'O'), ('1', 'O'), ('.', 'O'), ('32', 'O'), ('mln', 'O'), ('tonnes', 'unit'), ('in', 'O'), ('calendar', 'O'), ('1987', 'year'), ('.', 'O'), ('up', 'O'), ('from', 'O'), ('1', 'O'), ('.', 'O'), ('30', 'O'), ('mln', 'O'), ('tonnes', 'unit'), ('in', 'O'), ('1986', 'year'), ('.', 'O')]
```

```
Named Entity Recognition (NER) : {'tonnes', 'unit'}, ('1987', 'year'), ('Philippines', 'country'), ('1986', 'year'), ('Indonesia', 'country'), ('COPRA', 'crop'), ('copra', 'crop')}
```

Named Entity Recognition-

```
Named Entity Recognition (NER) : {'tonnes', 'unit'}, ('1987', 'year'), ('Philippines', 'country'), ('1986', 'year'), ('Indonesia', 'country'), ('COPRA', 'crop'), ('copra', 'crop')}
```

Measured Entity Recognition-

```
Measured Entity Detection: {'30,000 tonnes', '1.32 mln tonnes', '1.30 mln tonnes', '31 pct'}
```

Output 1 for “reuters/training/267” shows the correct result.

1.2 Test Case 2

Input 2- Reuter’s text training/279

JAPAN'S NTT FORECASTS PROFITS FALL IN

1987/88 <Nippon Telegraph and Telephone Corp> (NTT) expects its profits to fall to 328 billion yen in the year ending March 31, 1988 from a projected 348 billion this year, the company said. Total sales for the same period are expected to rise to 5,506 billion yen from a projected 5,328 billion this year, NTT said in a business operations plan submitted to the Post and Telecommunications Ministry. NTT said it plans to make capital investments of 1,770 billion yen in 1987/88, including 109 billion for research and development, as against a total of 1,600 billion this year. An NTT spokesman said increased competition from new entrants to the

telecommunications field and the effect of a sales tax scheduled to be introduced next January, were the major factors behind the projected decrease in profits. The Japanese telecommunications industry was deregulated in 1985.

Output 2-

Tokenization-

```
Tokens: ["JAPAN'S", 'NTT', 'FORECASTS', 'PROFITS', 'FALL', 'IN', '1987', '/', '88', '&', 'lt', ';', 'Nippon', 'Telegraph', 'and', 'Telephone', 'Corp', '>', '(', 'NTT', ')', 'expects', 'its', 'profits', 'to', 'fall', 'to', '328', 'billion', 'yen', 'in', 'the', 'year', 'ending', 'March', '31', ',', '1988', 'from', 'a', 'projected', '348', 'billion', 'this', 'year', ',', 'the', 'company', 'said', '.', 'Total', 'sales', 'for', 'the', 'same', 'period', 'are', 'expected', 'to', 'rise', 'to', '5', ',', '586', 'billion', 'yen', 'from', 'a', 'projected', '5', ',', '328', 'billion', 'this', 'year', ',', 'NTT', 'said', 'in', 'a', 'business', 'operations', 'plan', 'submitted', 'to', 'the', 'Post', 'and', 'Telecommunications', 'Ministry', '.', 'NTT', 'said', 'it', 'plans', 'to', 'make', 'capital', 'investments', 'of', '1', ',', '770', 'billion', 'yen', 'in', '1987', '/', '88', ',', 'including', '109', 'billion', 'for', 'research', 'and', 'development', ',', 'as', 'against', 'a', 'total', 'of', '1', ',', '600', 'billion', 'this', 'year', '.', 'An', 'NTT', 'spokesman', 'said', 'increased', 'competition', 'from', 'new', 'entrants', 'to', 'the', 'telecommunications', 'field', 'and', 'the', 'effect', 'of', 'a', 'sales', 'tax', 'scheduled', 'to', 'be', 'introduced', 'next', 'January', ',', 'were', 'the', 'major', 'factors', 'behind', 'the', 'projected', 'decrease', 'in', 'profits', '.', 'The', 'Japanese', 'telecommunications', 'industry', 'was', 'deregulated', 'in', '1985', '.']
```

Sentence Splitting-

```
Sentences: ["JAPAN'S NTT FORECASTS PROFITS FALL IN 1987/88\n &lt;Nippon Telegraph and Telephone Corp>\n (NTT) expects its profits to fall to 328 billion yen in the\n year ending March 31, 1988 from a projected 348 billion this\n year, the company said.", 'Total sales for the same period are expected to rise to\n 5,586 billion yen from a projected 5,328 billion this year, NTT\n said in a business operations plan submitted to the Post and\n Telecommunications Ministry.', 'NTT said it plans to make capital investments of 1,770\n billion yen in 1987/88, including 109 billion for research and\n development, as against a total of 1,600 billion this year.', 'An NTT spokesman said increased competition from new\n entrants to the telecommunications field and the effect of a\n sales tax scheduled to be introduced next January, were the\n major factors behind the projected decrease in profits.', 'The Japanese telecommunications industry was deregulated in\n 1985.']
```

POS Tagging-

```
POS Tags: [("JAPAN'S", 'NNP'), ('NTT', 'NNP'), ('FORECASTS', 'NNP'), ('PROFITS', 'NNP'), ('FALL', 'NNP'), ('IN', 'NNP'), ('1987', 'CD'), ('/', 'NNP'), ('88', 'CD'), ('&', 'CC'), ('lt', 'NN'), (';', ':'), ('Nippon', 'NNP'), ('Telegraph', 'NNP'), ('and', 'CC'), ('Telephone', 'NNP'), ('Corp', 'NNP'), ('>', 'NNP'), ('(', ' '), ('NTT', 'NNP'), (')', ' '), ('expects', 'VBZ'), ('its', 'PRP$'), ('profits', 'NNS'), ('to', 'TO'), ('fall', 'VB'), ('to', 'TO'), ('328', 'CD'), ('billion', 'CD'), ('yen', 'NNS'), ('in', 'IN'), ('the', 'DT'), ('year', 'NN'), ('ending', 'VBG'), ('March', 'NNP'), ('31', 'CD'), (',', ','), ('1988', 'CD'), ('from', 'IN'), ('a', 'DT'), ('projected', 'VBN'), ('348', 'CD'), ('billion', 'CD'), ('this', 'DT'), ('year', 'NN'), (',', ','), ('the', 'DT'), ('company', 'NN'), ('said', 'VBD'), ('.', ' '), ('Total', 'JJ'), ('sales', 'NNS'), ('for', 'IN'), ('the', 'DT'), ('same', 'JJ'), ('period', 'NN'), ('are', 'VBP'), ('expected', 'VBN'), ('to', 'TO'), ('rise', 'VB'), ('to', 'TO'), ('5', 'CD'), (',', ','), ('586', 'CD'), ('billion', 'CD'), ('yen', 'NNS'), ('from', 'IN'), ('a', 'DT'), ('projected', 'JJ'), ('5', 'CD'), (',', ','), ('328', 'CD'), ('billion', 'CD'), ('this', 'DT'), ('year', 'NN'), (',', ','), ('NTT', 'NNP'), ('said', 'VBD'), ('in', 'IN'), ('a', 'DT'), ('business', 'NN'), ('operations', 'NNS'), ('plan', 'NN'), ('submitted', 'VBN'), ('to', 'TO'), ('the', 'DT'), ('Post', 'NNP'), ('and', 'CC'), ('Telecommunications', 'NNP'), ('Ministry', 'NNP'), ('.', ' '), ('NTT', 'NNP'), ('said', 'VBD'), ('it', 'PRP'), ('plans', 'VBZ'), ('to', 'TO'), ('make', 'VB'), ('capital', 'NN'), ('investments', 'NNS'), ('of', 'IN'), ('1', 'CD'), (',', ','), ('770', 'CD'), ('billion', 'CD'), ('yen', 'NNS'), ('in', 'IN'), ('1987', 'CD'), ('/', 'NN'), ('88', 'CD'), (',', ','), ('including', 'VBG'), ('109', 'CD'), ('billion', 'CD'), ('for', 'IN'), ('research', 'NN'), ('and', 'CC'), ('development', 'NN'), (',', ','), ('as', 'IN'), ('against', 'IN'), ('a', 'DT'), ('total', 'NN'), ('of', 'IN'), ('1', 'CD'), (',', ','), ('600', 'CD'), ('billion', 'CD'), ('this', 'DT'), ('year', 'NN'), ('.', ' '), ('An', 'DT'), ('NTT', 'NNP'), ('spokesman', 'NN'), ('said', 'VBD'), ('increased', 'VBN'), ('competition', 'NN'), ('from', 'IN'), ('new', 'JJ'), ('entrants', 'NNS'), ('to', 'TO'), ('the', 'DT'), ('telecommunications', 'NNS'), ('field', 'NN'), ('and', 'CC'), ('the', 'DT'), ('effect', 'NN'), ('of', 'IN'), ('a', 'DT'), ('sales', 'NNS'), ('tax', 'NN'), ('scheduled', 'VBN'), ('to', 'TO'), ('be', 'VB'), ('introduced', 'VBN'), ('next', 'JJ'), ('January', 'NNP'), (',', ','), ('were', 'VBD'), ('the', 'DT'), ('major', 'JJ'), ('factors', 'NNS'), ('behind', 'IN'), ('the', 'DT'), ('projected', 'JJ'), ('decrease', 'NN'), ('in', 'IN'), ('profits', 'NNS'), ('.', ' '), ('The', 'DT'), ('Japanese', 'JJ'), ('telecommunications', 'NNS'), ('industry', 'NN'), ('was', 'VBD'), ('deregulated', 'VBN'), ('in', 'IN'), ('1985', 'CD'), ('.', ' ').]
```

Gazetteer Annotation-

```

Annotated Tokens: [{"JAPAN'S", '0'}, ('NTT', '0'), ('FORECASTS', '0'), ('PROFITS', '0'), ('FALL', '0'), ('IN', '0'), ('1987', 'year'), ('/', '0'), ('88', '0'), ('&', '0'), ('lt', '0'), (';', '0'), ('Nippon', '0'), ('Telegraph', '0'), ('and', '0'), ('Telephone', '0'), ('Corp', '0'), ('>', '0'), ('(', '0'), ('NTT', '0'), (')', '0'), ('expects', '0'), ('its', '0'), ('profits', '0'), ('to', '0'), ('fall', '0'), ('to', '0'), ('328', '0'), ('billion', '0'), ('yen', '0'), ('in', '0'), ('the', '0'), ('year', '0'), ('ending', '0'), ('March', '0'), ('31', '0'), (';', '0'), ('1988', 'year'), ('from', '0'), ('a', '0'), ('projected', '0'), ('348', '0'), ('billion', '0'), ('this', '0'), ('year', '0'), (';', '0'), ('the', '0'), ('company', '0'), ('said', '0'), (';', '0'), ('Total', '0'), ('sales', '0'), ('for', '0'), ('the', '0'), ('same', '0'), ('period', '0'), ('are', '0'), ('expected', '0'), ('to', '0'), ('rise', '0'), ('to', '0'), ('5', '0'), (';', '0'), ('586', '0'), ('billion', '0'), ('yen', '0'), ('from', '0'), ('a', '0'), ('projected', '0'), ('5', '0'), (';', '0'), ('328', '0'), ('billion', '0'), ('this', '0'), ('year', '0'), (';', '0'), ('NTT', '0'), ('said', '0'), ('in', '0'), ('a', '0'), ('business', '0'), ('operations', '0'), ('plan', '0'), ('submitted', '0'), ('to', '0'), ('the', '0'), ('Post', '0'), ('and', '0'), ('Telecommunications', '0'), ('Ministry', '0'), ('.', '0'), ('NTT', '0'), ('said', '0'), ('it', '0'), ('plans', '0'), ('to', '0'), ('make', '0'), ('capital', '0'), ('investments', '0'), ('of', '0'), ('1', '0'), (';', '0'), ('778', '0'), ('billion', '0'), ('yen', '0'), ('in', '0'), ('1987', 'year'), ('/', '0'), ('88', '0'), (';', '0'), ('including', '0'), ('189', '0'), ('billion', '0'), ('for', '0'), ('research', '0'), ('and', '0'), ('development', '0'), (';', '0'), ('as', '0'), ('against', '0'), ('a', '0'), ('total', '0'), ('of', '0'), ('1', '0'), (';', '0'), ('688', '0'), ('billion', '0'), ('this', '0'), ('year', '0'), ('.', '0'), ('An', '0'), ('NTT', '0'), ('spokesman', '0'), ('said', '0'), ('increased', '0'), ('competition', '0'), ('from', '0'), ('new', '0'), ('entrants', '0'), ('to', '0'), ('the', '0'), ('telecommunications', '0'), ('field', '0'), ('and', '0'), ('the', '0'), ('effect', '0'), ('of', '0'), ('a', '0'), ('sales', '0'), ('tax', '0'), ('scheduled', '0'), ('to', '0'), ('be', '0'), ('introduced', '0'), ('next', '0'), ('January', '0'), (';', '0'), ('were', '0'), ('the', '0'), ('major', '0'), ('factors', '0'), ('behind', '0'), ('the', '0'), ('projected', '0'), ('decrease', '0'), ('in', '0'), ('profits', '0'), ('.', '0'), ('The', '0'), ('Japanese', '0'), ('telecommunications', '0'), ('industry', '0'), ('was', '0'), ('deregulated', '0'), ('in', '0'), ('1985', 'year'), ('.', '0')]

```

Named Entity Recognition-

```

Named Entity Recognition (NER) : {'1985', 'year'}, ('1988', 'year'), ('JAPAN', 'country')}

```

Measured Entity Recognition-

```

Measured Entity Detection: {'5,328 billion', '328 billion', '189 billion', '5,586 billion', '1,688 billion', '348 billion', '1,778 billion'}

```

Output 2 for “reuters/training/267”, 1987/88 is not predicted as year.

1.3 Test Case 3

Input 3- Reuter’s text training/9865

FRENCH FREE MARKET CEREAL EXPORT BIDS DETAILED

French operators have requested licences to export 675,500 tonnes of maize, 245,000 tonnes of barley, 22,000 tonnes of soft bread wheat and 20,000 tonnes of feed wheat at today's European Community tender, traders said. Rebates requested ranged from 127.75 to 132.50 European Currency Units a tonne for maize, 136.00 to 141.00 Ecus a tonne for barley and 134.25 to 141.81 Ecus for bread wheat, while rebates requested for feed wheat were 137.65 Ecus, they said.

Output 3-

Tokenization-

```

Tokens: ['FRENCH', 'FREE', 'MARKET', 'CEREAL', 'EXPORT', 'BIDS', 'DETAILED', 'French', 'operators', 'have', 'requested', 'licences', 'to', 'export', '675', ',', '500', 'tonnes', 'of', 'maize', ',', '245', ',', '000', 'tonnes', 'of', 'barley', ',', '22', ',', '000', 'tonnes', 'of', 'soft', 'bread', 'wheat', 'and', '20', ',', '000', 'tonnes', 'of', 'feed', 'wheat', 'at', "today's", 'European', 'Community', 'tender', ',', 'traders', 'said', '.,', 'Rebates', 'requested', 'ranged', 'from', '127', '.,', '75', 'to', '132', '.,', '50', 'European', 'Currency', 'Units', 'a', 'tonne', 'for', 'maize', ',', '136', '.,', '80', 'to', '141', '.,', '80', 'Ecus', 'a', 'tonne', 'for', 'barley', 'and', '134', '.,', '25', 'to', '141', '.,', '81', 'Ecus', 'for', 'bread', 'wheat', ',', 'while', 'rebates', 'requested', 'for', 'feed', 'wheat', 'were', '137', '.,', '65', 'Ecus', ',', 'they', 'said', '.,']

```

Sentence Splitting-

```
Sentences: ["FRENCH FREE MARKET CEREAL EXPORT BIDS DETAILED\n French operators have requested licences\n to export 675,500 tonnes of maize, 245,000 tonnes of  
barley,\n 22,000 tonnes of soft bread wheat and 20,000 tonnes of feed\n wheat at today's European Community tender, traders said.", 'Rebates requested ranged from  
127.75 to 132.50 European\n Currency Units a tonne for maize, 136.00 to 141.00 Ecus a tonne\n for barley and 134.25 to 141.81 Ecus for bread wheat, while\n rebates requested for feed wheat were 137.65 Ecus, they said.']
```

POS Tagging-

```
POS Tags: [('FRENCH', 'JJ'), ('FREE', 'NNP'), ('MARKET', 'NNP'), ('CEREAL', 'NNP'), ('EXPORT', 'NNP'), ('BIDS', 'NNP'), ('DETAILED', 'NNP'), ('French', 'JJ'),  
(('operators', 'NNS'), ('have', 'VBP'), ('requested', 'VBN'), ('licences', 'NNS'), ('to', 'TO'), ('export', 'VB'), ('675', 'CD'), (',', ','), ('500', 'CD'),  
(('tonnes', 'NNS'), ('of', 'IN'), ('maize', 'NN'), ('.', '.'), ('245', 'CD'), (',', ','), ('800', 'CD'), ('tonnes', 'NNS'), ('of', 'IN'), ('barley', 'NN'), (',', ','),  
(('22', 'CD'), (',', ','), ('800', 'CD'), ('tonnes', 'NNS'), ('of', 'IN'), ('soft', 'JJ'), ('bread', 'NN'), ('wheat', 'NN'), ('and', 'CC'), ('20', 'CD'), (',', ','),  
(('000', 'CD'), ('tonnes', 'NNS'), ('of', 'IN'), ('feed', 'NN'), ('wheat', 'NN'), ('at', 'IN'), ('today's', 'JJ'), ('European', 'JJ'), ('Community', 'NNP'),  
(('tender', 'NN'), (',', ','), ('traders', 'NNS'), ('said', 'VBD'), (',', '.'), ('Rebates', 'NNS'), ('requested', 'VBD'), ('ranged', 'VBD'), ('from', 'IN'), ('127',  
(('CD'), (',', '.'), ('75', 'CD'), ('to', 'TO'), ('132', 'CD'), (',', '.'), ('50', 'CD'), ('European', 'JJ'), ('Currency', 'NNP'), ('Units', 'VBZ'), ('a', 'DT'),  
(('tonne', 'NN'), ('for', 'IN'), ('maize', 'NN'), (',', ','), ('136', 'CD'), (',', '.'), ('00', 'CD'), ('to', 'TO'), ('141', 'CD'), (',', '.'), ('00', 'CD'),  
(('Ecus', 'NNP'), ('a', 'DT'), ('tonne', 'NN'), ('for', 'IN'), ('barley', 'NN'), ('and', 'CC'), ('134', 'CD'), (',', '.'), ('25', 'CD'), ('to', 'TO'), ('141', 'CD'),  
(',', '.'), ('81', 'CD'), ('Ecus', 'NNP'), ('for', 'IN'), ('bread', 'JJ'), ('wheat', 'NN'), (',', ','), ('while', 'IN'), ('rebates', 'NNS'), ('requested', 'VBD'),  
(('for', 'IN'), ('feed', 'NN'), ('wheat', 'NN'), ('were', 'VBD'), ('137', 'CD'), (',', '.'), ('65', 'CD'), ('Ecus', 'NNP'), (',', ','), ('they', 'PRP'), ('said',  
(('VBD'), (',', '.'))]
```

Gazetteer Annotation-

```
Annotated Tokens: [('FRENCH', 'O'), ('FREE', 'O'), ('MARKET', 'O'), ('CEREAL', 'O'), ('EXPORT', 'O'), ('BIDS', 'O'), ('DETAILED', 'O'), ('French', 'O'),  
(('operators', 'O'), ('have', 'O'), ('requested', 'O'), ('licences', 'O'), ('to', 'O'), ('export', 'O'), ('675', 'O'), (',', 'O'), ('500', 'O'), ('tonnes', 'unit'),  
(('of', 'O'), ('maize', 'O'), (',', 'O'), ('245', 'O'), (',', 'O'), ('800', 'O'), ('tonnes', 'unit'), ('of', 'O'), ('barley', 'O'), (',', 'O'), ('22', 'O'), (',',  
'O'), ('800', 'O'), ('tonnes', 'unit'), ('of', 'O'), ('soft', 'O'), ('bread', 'O'), ('wheat', 'crop'), ('and', 'O'), ('20', 'O'), (',', 'O'), ('000', 'O'),  
(('tonnes', 'unit'), ('of', 'O'), ('feed', 'O'), ('wheat', 'crop'), ('at', 'O'), ('today's', 'O'), ('European', 'O'), ('Community', 'O'), ('tender', 'O'), (',',  
'O'), ('traders', 'O'), ('said', 'O'), (',', 'O'), ('Rebates', 'O'), ('requested', 'O'), ('ranged', 'O'), ('from', 'O'), ('127', 'O'), (',', 'O'), ('75', 'O'),  
(('to', 'O'), ('132', 'O'), (',', 'O'), ('50', 'O'), ('European', 'O'), ('Currency', 'O'), ('Units', 'O'), ('a', 'O'), ('tonne', 'O'), ('for', 'O'), ('maize', 'O'),  
(',', 'O'), ('136', 'O'), (',', 'O'), ('00', 'O'), ('to', 'O'), ('141', 'O'), (',', 'O'), ('00', 'O'), ('Ecus', 'O'), ('a', 'O'), ('tonne', 'O'), ('for', 'O'),  
(('barley', 'O'), ('and', 'O'), ('134', 'O'), (',', 'O'), ('25', 'O'), ('to', 'O'), ('141', 'O'), (',', 'O'), ('81', 'O'), ('Ecus', 'O'), ('for', 'O'), ('bread',  
'O'), ('wheat', 'crop'), (',', 'O'), ('while', 'O'), ('rebates', 'O'), ('requested', 'O'), ('for', 'O'), ('feed', 'O'), ('wheat', 'crop'), ('were', 'O'), ('137',  
'O'), (',', 'O'), ('65', 'O'), ('Ecus', 'O'), (',', 'O'), ('they', 'O'), ('said', 'O'), (',', 'O')]
```

Named Entity Recognition-

```
Named Entity Recognition (NER) : {('wheat', 'crop'), ('tonnes', 'unit')}
```

Measured Entity Recognition-

```
Measured Entity Detection: {'22,000 tonnes', '141.00 Ecus', '137.65 Ecus', '20,000 tonnes', '245,000 tonnes', '141.81 Ecus', '675,500 tonnes'}
```

For “reuters/training/9865”, the output shows the correct result.

1.4 Test Case 4

Input 4- Reuter’s text training/9880

U.K., MONEY MARKET GETS 25 MLN STG LATE HELP

The Bank of England said it provided about 25 mln stg in late help to the money market, bringing the total assistance today to 266 mln stg. This compares with the bank's revised estimate of a 350 mln stg money market shortfall.

Output 4-

Tokenization-

```
Tokens: ['U', '.', 'K', '.', 'MONEY', 'MARKET', 'GETS', '25', 'MLN', 'STG', 'LATE', 'HELP', 'The', 'Bank', 'of', 'England', 'said', 'it', 'provided', 'about', '25', 'mln', 'stg', 'in', 'late', 'help', 'to', 'the', 'money', 'market', ',', 'bringing', 'the', 'total', 'assistance', 'today', 'to', '266', 'mln', 'stg', '.', 'This', 'compares', 'with', 'the', 'bank's', 'revised', 'estimate', 'of', 'a', '350', 'mln', 'stg', 'money', 'market', 'shortfall', '.']
```

Sentence Splitting-

```
Sentences: ['U.K.', 'MONEY MARKET GETS 25 MLN STG LATE HELP\n The Bank of England said it provided\n about 25 mln stg in late help to the money market, bringing the\n total assistance today to 266 mln stg.', 'This compares with the bank's revised estimate of a 350 mln\n stg money market shortfall.']
```

POS Tagging-

```
POS Tags: [('U', 'NNP'), ('.', '.'), ('K', 'NNP'), ('.', '.'), ('MONEY', 'NN'), ('MARKET', 'NNP'), ('GETS', 'NNP'), ('25', 'CD'), ('MLN', 'NNP'), ('STG', 'NNP'), ('LATE', 'NNP'), ('HELP', 'NNP'), ('The', 'DT'), ('Bank', 'NNP'), ('of', 'IN'), ('England', 'NNP'), ('said', 'VBD'), ('it', 'PRP'), ('provided', 'VBD'), ('about', 'RB'), ('25', 'CD'), ('mln', 'NNS'), ('stg', 'RB'), ('in', 'IN'), ('late', 'JJ'), ('help', 'NN'), ('to', 'TO'), ('the', 'DT'), ('money', 'NN'), ('market', 'NN'), (',', ','), ('bringing', 'VBG'), ('the', 'DT'), ('total', 'JJ'), ('assistance', 'NN'), ('today', 'NN'), ('to', 'TO'), ('266', 'CD'), ('mln', 'NN'), ('stg', 'NN'), ('.', '.'), ('This', 'DT'), ('compares', 'VBZ'), ('with', 'IN'), ('the', 'DT'), ('bank's', 'NN'), ('revised', 'VBD'), ('estimate', 'NN'), ('of', 'IN'), ('a', 'DT'), ('350', 'CD'), ('mln', 'NN'), ('stg', 'JJ'), ('money', 'NN'), ('market', 'NN'), ('shortfall', 'NN'), ('.', '.')] 
```

Gazetteer Annotation-

```
Annotated Tokens: [('U', 'O'), ('.', 'O'), ('K', 'O'), ('.', 'O'), ('MONEY', 'O'), ('MARKET', 'O'), ('GETS', 'O'), ('25', 'O'), ('MLN', 'O'), ('STG', 'O'), ('LATE', 'O'), ('HELP', 'O'), ('The', 'O'), ('Bank', 'O'), ('of', 'O'), ('England', 'country'), ('said', 'O'), ('it', 'O'), ('provided', 'O'), ('about', 'O'), ('25', 'O'), ('mln', 'O'), ('stg', 'O'), ('in', 'O'), ('late', 'O'), ('help', 'O'), ('to', 'O'), ('the', 'O'), ('money', 'O'), ('market', 'O'), (',', 'O'), ('bringing', 'O'), ('the', 'O'), ('total', 'O'), ('assistance', 'O'), ('today', 'O'), ('to', 'O'), ('266', 'O'), ('mln', 'O'), ('stg', 'O'), ('.', 'O'), ('This', 'O'), ('compares', 'O'), ('with', 'O'), ('the', 'O'), ('bank's', 'O'), ('revised', 'O'), ('estimate', 'O'), ('of', 'O'), ('a', 'O'), ('350', 'O'), ('mln', 'O'), ('stg', 'O'), ('money', 'O'), ('market', 'O'), ('shortfall', 'O'), ('.', 'O')]
```

Named Entity Recognition-

```
Named Entity Recognition (NER) : {('U.K.', 'country'), ('England', 'country')}
```

Measured Entity Recognition-

```
Measured Entity Detection: {'25 mln stg', '266 mln stg', '25 MLN STG'}
```

For “reuters/training/9880”, the sentence is splitting from the country “U.K.” that is not correct. The rest modules shows the correct result.

2. Interesting Case

2.1 Test Case 5

Input 5- Reuter's text training/6

ARGENTINE 1986/87 GRAIN/OILSEED REGISTRATIONS

Argentine grain board figures show crop registrations of grains, oilseeds and their products to February 11, in thousands of tonnes, showing those for futurE shipments month, 1986/87 total and 1985/86 total to February 12, 1986, in brackets:

Bread wheat prev 1,655.8, Feb 872.0, March 164.6, total 2,692.4 (4,161.0).

Maize Mar 48.0, total 48.0 (nil).

Sorghum nil (nil)

Oilseed export registrations were:

Sunflowerseed total 15.0 (7.9)

Soybean May 20.0, total 20.0 (nil)

The board also detailed export registrations for subproducts, as follows,

SUBPRODUCTS

Wheat prev 39.9, Feb 48.7, March 13.2, Apr 10.0, total 111.8 (82.7) .

Linseed prev 34.8, Feb 32.9, Mar 6.8, Apr 6.3, total 80.8 (87.4).

Soybean prev 100.9, Feb 45.1, MAr nil, Apr nil, May 20.0, total 166.1 (218.5).

Sunflowerseed prev 48.6, Feb 61.5, Mar 25.1, Apr 14.5, total 149.8 (145.3).

Vegetable oil registrations were :

Sunoil prev 37.4, Feb 107.3, Mar 24.5, Apr 3.2, May nil, Jun 10.0, total 182.4 (117.6).

Linoil prev 15.9, Feb 23.6, Mar 20.4, Apr 2.0, total 61.8, (76.1).

Soybean oil prev 3.7, Feb 21.1, Mar nil, Apr 2.0, May 9.0, Jun 13.0, Jul 7.0, total 55.8 (33.7).

REUTER

Output 5-

Tokenization-

```
Tokens: ['ARGENTINE', '1986', '/', '87', 'GRAIN', '/', 'OILSEED', 'REGISTRATIONS', 'Argentine', 'grain', 'board', 'figures', 'show', 'crop', 'registrations', 'of', 'grains', ',', 'oilseeds', 'and', 'their', 'products', 'to', 'February', '11', ',', 'in', 'thousands', 'of', 'tonnes', ',', 'showing', 'those', 'for', 'futurE', 'shipments', 'month', ',', '1986', '/', '87', 'total', 'and', '1985', '/', '86', 'total', 'to', 'February', '12', ',', '1986', ',', 'in', 'brackets', ':', 'Bread', 'wheat', 'prev', '1', ',', '655', ',', '8', ',', 'Feb', '872', ',', '0', ',', 'March', '164', ',', '6', ',', 'total', '2', ',', '692', ',', '4', '(', '4', ',', '161', ',', '0', ')', ',', 'Maize', 'Mar', '48', ',', '0', ',', 'total', '48', ',', '0', '(', 'nil', ')', ',', 'Sorghum', 'nil', '(', 'nil', ')', 'Oilseed', 'export', 'registrations', 'were', ':', 'Sunflowerseed', 'total', '15', ',', '0', '(', '7', ',', '9', ')', 'Soybean', 'May', '20', ',', '0', ',', 'total', '20', ',', '0', '(', 'nil', ')', 'The', 'board', 'also', 'detailed', 'export', 'registrations', 'for', 'subproducts', ',', 'as', 'follows', ',', 'SUBPRODUCTS', 'Wheat', 'prev', '39', ',', '9', ',', 'Feb', '48', ',', '7', ',', 'March', '13', ',', '2', ',', 'Apr', '10', ',', '0', ',', 'total', '111', ',', '8', '(', '82', ',', '7', ')', ',', 'Linseed', 'prev', '34', ',', '8', ',', 'Feb', '32', ',', '9', ',', 'Mar', '6', ',', '8', ',', 'Apr', '6', ',', '3', ',', 'total', '80', ',', '8', '(', '87', ',', '4', ')', ',', 'Soybean', 'prev', '100', ',', '9', ',', 'Feb', '45', ',', '1', ',', 'MAr', 'nil', ',', 'Apr', 'nil', ',', 'May', '20', ',', '0', ',', 'total', '166', ',', '1', '(', '218', ',', '5', ')', ',', 'Sunflowerseed', 'prev', '48', ',', '6', ',', 'Feb', '61', ',', '5', ',', 'Mar', '25', ',', '1', ',', 'Apr', '14', ',', '5', ',', 'total', '149', ',', '8', '(', '145', ',', '3', ')', ',', 'Vegetable', 'oil', 'registrations', 'were', ':', 'Sunoil', 'prev', '37', ',', '4', ',', 'Feb', '107', ',', '3', ',', 'Mar', '24', ',', '5', ',', 'Apr', '3', ',', '2', ',', 'May', 'nil', ',', 'Jun', '10', ',', '0', ',', 'total', '182', ',', '4', '(', '117', ',', '6', ')', ',', 'Linoil', 'prev', '15', ',', '9', ',', 'Feb', '23', ',', '6', ',', 'Mar', '20', ',', '4', ',', 'Apr', '2', ',', '0', ',', 'total', '61', ',', '8', '(', '76', ',', '1', ')', ',', 'Soybean', 'oil', 'prev', '3', ',', '7', ',', 'Feb', '21', ',', '1', ',', 'Mar', 'nil', ',', 'Apr', '2', ',', '0', ',', 'May', '9', ',', '0', ',', 'Jun', '13', ',', '0', ',', 'Jul', '7', ',', '0', ',', 'total', '55', ',', '8', '(', '33', ',', '7', ')', ',', 'REUTER']
```

Sentence Splitting-

Sentences: ['ARGENTINE 1986/87 GRAIN/OILSEED REGISTRATIONS\n Argentine grain board figures show\n crop registrations of grains, oilseeds and their products to\n February 11, in thousands of tonnes, showing those for futurE\n shipments month, 1986/87 total and 1985/86 total to February\n 12, 1986, in brackets:\n Bread wheat prev 1,655.8, Feb 872.0, March 164.6, total\n 2,692.4 (4,161.0).', 'Maize Mar 48.0, total 48.0 (nil).', 'Sorghum nil (nil)\n Oilseed export registrations were:\n Sunflowerseed total 15.0 (7.9)\n Soybean May 20.0, total 20.0 (nil)\n The board also detailed export registrations for\n subproducts, as follows,\n SUBPRODUCTS\n Wheat prev 39.9, Feb 48.7, March 13.2, Apr 10.0, total\n 111.8 (82.7) .', 'Linseed prev 34.8, Feb 32.9, Mar 6.8, Apr 6.3, total 80.8\n (87.4).', 'Soybean prev 100.9, Feb 45.1, MAR nil, Apr nil, May 20.0,\n total 166.1 (218.5).', 'Sunflowerseed prev 48.6, Feb 61.5, Mar 25.1, Apr 14.5,\n total 149.8 (145.3).', 'Vegetable oil registrations were : \n Sunoil prev 37.4, Feb 107.3, Mar 24.5, Apr 3.2, May nil,\n Jun 10.0, total 182.4 (117.6).', 'Linoil prev 15.9, Feb 23.6, Mar 20.4, Apr 2.0, total 61.8,\n (76.1).', 'Soybean oil prev 3.7, Feb 21.1, Mar nil, Apr 2.0, May 9.0,\n Jun 13.0, Jul 7.0, total 55.8 (33.7).', 'REUTER']

POS Tagging-

POS Tags: [['ARGENTINE', 'NNP'], ('1986', 'CD'), ('/', 'NNP'), ('87', 'CD'), ('GRAIN', 'NNP'), ('/', 'NNP'), ('OILSEED', 'NNP'), ('REGISTRATIONS', 'NNP'), ('Argentine', 'NNP'), ('grain', 'NN'), ('board', 'NN'), ('figures', 'VBZ'), ('show', 'VBP'), ('crop', 'NN'), ('registrations', 'NNS'), ('of', 'IN'), ('grains', 'NNS'), ('', ' ', ''), ('oilseeds', 'NNS'), ('and', 'CC'), ('their', 'PRP\$'), ('products', 'NNS'), ('to', 'TO'), ('February', 'NNP'), ('11', 'CD'), ('', ' ', ''), ('in', 'IN'), ('thousands', 'NNS'), ('of', 'IN'), ('tonnes', 'NNS'), ('', ' ', ''), ('showing', 'VBG'), ('those', 'DT'), ('for', 'IN'), ('future', 'JJ'), ('shipments', 'NNS'), ('month', 'NN'), ('', ' ', ''), ('1986', 'CD'), ('/', 'VBD'), ('87', 'CD'), ('total', 'NN'), ('and', 'CC'), ('1985', 'CD'), ('/', 'NN'), ('86', 'CD'), ('total', 'NN'), ('to', 'TO'), ('February', 'NNP'), ('12', 'CD'), ('', ' ', ''), ('1986', 'CD'), ('', ' ', ''), ('in', 'IN'), ('brackets', 'NNS'), (':', ':'), ('Bread', 'NNP'), ('wheat', 'NN'), ('prev', 'NN'), ('1', 'CD'), ('', ' ', ''), ('655', 'CD'), ('', ' ', ''), ('8', 'CD'), ('', ' ', ''), ('Feb', 'NNP'), ('872', 'CD'), ('', ' ', ''), ('0', 'CD'), ('', ' ', ''), ('March', 'NNP'), ('164', 'CD'), ('', ' ', ''), ('6', 'CD'), ('', ' ', ''), ('total', 'JJ'), ('2', 'CD'), ('', ' ', ''), ('692', 'CD'), ('', ' ', ''), ('4', 'CD'), ('', ' ', ''), ('4', 'CD'), ('', ' ', ''), ('161', 'CD'), ('', ' ', ''), ('0', 'CD'), ('', ' ', ''), ('', ' ', ''), ('Maize', 'NNP'), ('Mar', 'NNP'), ('48', 'CD'), ('', ' ', ''), ('0', 'CD'), ('', ' ', ''), ('total', 'JJ'), ('48', 'CD'), ('', ' ', ''), ('0', 'CD'), ('(', '(', 'nil', 'NN'), ('', ' ', ''), ('', ' ', ''), ('Sorghum', 'NNP'), ('nil', 'NN'), ('(', '(', 'nil', 'JJ'), ('', ' ', ''), ('oilseed', 'NNP'), ('export', 'NN'), ('registrations', 'NNS'), ('were', 'VBD'), (':', ':'), ('Sunflowerseed', 'NNP'), ('total', 'VBP'), ('15', 'CD'), ('', ' ', ''), ('0', 'CD'), ('(', '(', '7', 'CD'), ('', ' ', ''), ('9', 'CD'), ('', ' ', ''), ('Soybean', 'NN'), ('May', 'NNP'), ('20', 'CD'), ('', ' ', ''), ('0', 'CD'), ('', ' ', ''), ('total', 'JJ'), ('20', 'CD'), ('', ' ', ''), ('0', 'CD'), ('(', '(', 'nil', 'NN'), ('', ' ', ''), ('The', 'DT'), ('board', 'NN'), ('also', 'RB'), ('detailed', 'VBD'), ('export', 'NN'), ('registrations', 'NNS'), ('for', 'IN'), ('subproducts', 'NNS'), ('', ' ', ''), ('as', 'IN'), ('follows', 'VBZ'), ('', ' ', ''), ('SUBPRODUCTS', 'NNP'), ('Wheat', 'NNP'), ('prev', 'VBD'), ('39', 'CD'), ('', ' ', ''), ('9', 'CD'), ('', ' ', ''), ('Feb', 'NNP'), ('48', 'CD'), ('', ' ', ''), ('7', 'CD'), ('', ' ', ''), ('March', 'NNP'), ('13', 'CD'), ('', ' ', ''), ('2', 'CD'), ('', ' ', ''), ('Apr', 'NNP'), ('10', 'CD'), ('', ' ', ''), ('0', 'CD'), ('', ' ', ''), ('total', 'JJ'), ('111', 'CD'), ('', ' ', ''), ('8', 'CD'), ('(', '(', '82', 'CD'), ('', ' ', ''), ('7', 'CD'), ('', ' ', ''), ('Linseed', 'NNP'), ('prev', 'VBZ'), ('34', 'CD'), ('', ' ', ''), ('8', 'CD'), ('', ' ', ''), ('Feb', 'NNP'), ('32', 'CD'), ('', ' ', ''), ('9', 'CD'), ('', ' ', ''), ('Mar', 'NNP'), ('6', 'CD'), ('', ' ', ''), ('8', 'CD'), ('', ' ', ''), ('Apr', 'NNP'), ('6', 'CD'), ('', ' ', ''), ('3', 'CD'), ('', ' ', ''), ('total', 'JJ'), ('80', 'CD'), ('', ' ', ''), ('8', 'CD'), ('(', '(', '87', 'CD'), ('', ' ', ''), ('4', 'CD'), ('', ' ', ''), ('', ' ', ''), ('Soybean', 'NNP'), ('prev', 'VBD'), ('100', 'CD'), ('', ' ', ''), ('9', 'CD'), ('', ' ', ''), ('Feb', 'NNP'), ('45', 'CD'), ('', ' ', ''), ('1', 'CD'), ('', ' ', ''), ('Mar', 'NNP'), ('nil', 'NN'), ('', ' ', ''), ('Apr', 'NNP'), ('nil', 'RB'), ('', ' ', ''), ('May', 'NNP'), ('20', 'CD'), ('', ' ', ''), ('0', 'CD'), ('', ' ', ''), ('total', 'JJ'), ('166', 'CD'), ('', ' ', ''), ('1', 'CD'), ('(', '(', '218', 'CD'), ('', ' ', ''), ('5', 'CD'), ('', ' ', ''), ('Sunflowerseed', 'NNP'), ('prev', 'VBZ'), ('48', 'CD'), ('', ' ', ''), ('6', 'CD'), ('', ' ', ''), ('Feb', 'NNP'), ('61', 'CD'), ('', ' ', ''), ('5', 'CD'), ('', ' ', ''), ('Mar', 'NNP'), ('25', 'CD'), ('', ' ', ''), ('1', 'CD'), ('', ' ', ''), ('Apr', 'NNP'), ('14', 'CD'), ('', ' ', ''), ('5', 'CD'), ('', ' ', ''), ('total', 'JJ'), ('149', 'CD'), ('', ' ', ''), ('8', 'CD'), ('(', '(', '145', 'CD'), ('', ' ', ''), ('3', 'CD'), ('', ' ', ''), ('Vegetable', 'JJ'), ('oil', 'NN'), ('registrations', 'NNS'), ('were', 'VBD'), (':', ':'), ('Sunoil', 'NNP'), ('prev', 'VBZ'), ('37', 'CD'), ('', ' ', ''), ('4', 'CD'), ('', ' ', ''), ('Feb', 'NNP'), ('107', 'CD'), ('', ' ', ''), ('3', 'CD'), ('', ' ', ''), ('Mar', 'NNP'), ('24', 'CD'), ('', ' ', ''), ('5', 'CD'), ('', ' ', ''), ('Apr', 'NNP'), ('3', 'CD'), ('', ' ', ''), ('2', 'CD'), ('', ' ', ''), ('May', 'NNP'), ('nil', 'NN'), ('', ' ', ''), ('Jun', 'NNP'), ('10', 'CD'), ('', ' ', ''), ('0', 'CD'), ('', ' ', ''), ('total', 'JJ'), ('182', 'CD'), ('', ' ', ''), ('4', 'CD'), ('(', '(', '117', 'CD'), ('', ' ', ''), ('6', 'CD'), ('', ' ', ''), ('Linoil', 'NNP'), ('prev', 'VBD'), ('15', 'CD'), ('', ' ', ''), ('9', 'CD'), ('', ' ', ''), ('Feb', 'NNP'), ('23', 'CD'), ('', ' ', ''), ('6', 'CD'), ('', ' ', ''), ('Mar', 'NNP'), ('20', 'CD'), ('', ' ', ''), ('4', 'CD'), ('', ' ', ''), ('Apr', 'NNP'), ('2', 'CD'), ('', ' ', ''), ('0', 'CD'), ('', ' ', ''), ('total', 'JJ'), ('61', 'CD'), ('', ' ', ''), ('8', 'CD'), ('(', '(', '76', 'CD'), ('', ' ', ''), ('1', 'CD'), ('', ' ', ''), ('Soybean', 'JJ'), ('oil', 'NN'), ('prev', 'NN'), ('3', 'CD'), ('', ' ', ''), ('7', 'CD'), ('', ' ', ''), ('Feb', 'NNP'), ('21', 'CD'), ('', ' ', ''), ('1', 'CD'), ('', ' ', ''), ('Mar', 'NNP'), ('nil', 'NN'), ('', ' ', ''), ('Apr', 'NNP'), ('2', 'CD'), ('', ' ', ''), ('0', 'CD'), ('', ' ', ''), ('May', 'NNP'), ('9', 'CD'), ('', ' ', ''), ('0', 'CD'), ('', ' ', ''), ('Jun', 'NNP'), ('13', 'CD'), ('', ' ', ''), ('0', 'CD'), ('', ' ', ''), ('Jul', 'NNP'), ('7', 'CD'), ('', ' ', ''), ('0', 'CD'), ('', ' ', ''), ('total', 'JJ'), ('55', 'CD'), ('', ' ', ''), ('8', 'CD'), ('(', '(', '33', 'CD'), ('', ' ', ''), ('7', 'CD'), ('', ' ', ''), ('REUTER', 'NN')]

Gazetteer Annotation-

Annotated Tokens: [('ARGENTINE', '0'), ('1986', 'year'), ('/', '0'), ('87', '0'), ('GRAIN', '0'), ('/', '0'), ('OILSEED', '0'), ('REGISTRATIONS', '0'), ('Argentine', '0'), ('grain', '0'), ('board', '0'), ('figures', '0'), ('show', '0'), ('crop', '0'), ('registrations', '0'), ('of', '0'), ('grains', '0'), ('', '0'), ('oilseeds', '0'), ('and', '0'), ('their', '0'), ('products', '0'), ('to', '0'), ('February', '0'), ('11', '0'), ('', '0'), ('in', '0'), ('thousands', '0'), ('of', '0'), ('tonnes', 'unit'), ('', '0'), ('showing', '0'), ('those', '0'), ('for', '0'), ('future', '0'), ('shipments', '0'), ('month', '0'), ('', '0'), ('1986', 'year'), ('/', '0'), ('87', '0'), ('total', '0'), ('and', '0'), ('1985', 'year'), ('/', '0'), ('86', '0'), ('total', '0'), ('to', '0'), ('February', '0'), ('12', '0'), ('', '0'), ('1986', 'year'), ('', '0'), ('in', '0'), ('brackets', '0'), (':', '0'), ('Bread', '0'), ('wheat', 'crop'), ('prev', '0'), ('1', '0'), ('', '0'), ('655', '0'), ('', '0'), ('8', '0'), ('', '0'), ('Feb', '0'), ('872', '0'), ('', '0'), ('8', '0'), ('', '0'), ('March', '0'), ('164', '0'), ('', '0'), ('6', '0'), ('', '0'), ('total', '0'), ('2', '0'), ('', '0'), ('692', '0'), ('', '0'), ('4', '0'), ('(', '0'), ('4', '0'), ('', '0'), ('161', '0'), ('', '0'), ('8', '0'), ('', '0'), ('', '0'), ('Maize', '0'), ('Mar', '0'), ('48', '0'), ('', '0'), ('8', '0'), ('', '0'), ('total', '0'), ('48', '0'), ('', '0'), ('8', '0'), ('(', '0'), ('nil', '0'), ('', '0'), ('', '0'), ('Sorghum', '0'), ('nil', '0'), ('(', '0'), ('nil', '0'), ('', '0'), ('oilseed', '0'), ('export', '0'), ('registrations', '0'), ('were', '0'), (':', '0'), ('Sunflowerseed', 'crop'), ('total', '0'), ('15', '0'), ('', '0'), ('8', '0'), ('(', '0'), ('7', '0'), ('', '0'), ('9', '0'), ('', '0'), ('Soybean', '0'), ('May', '0'), ('28', '0'), ('', '0'), ('8', '0'), ('', '0'), ('total', '0'), ('28', '0'), ('', '0'), ('8', '0'), ('(', '0'), ('nil', '0'), ('', '0'), ('The', '0'), ('board', '0'), ('also', '0'), ('detailed', '0'), ('export', '0'), ('registrations', '0'), ('for', '0'), ('subproducts', '0'), ('', '0'), ('as', '0'), ('follows', '0'), ('', '0'), ('SUBPRODUCTS', '0'), ('Wheat', '0'), ('prev', '0'), ('39', '0'), ('', '0'), ('9', '0'), ('', '0'), ('Feb', '0'), ('48', '0'), ('', '0'), ('7', '0'), ('', '0'), ('March', '0'), ('13', '0'), ('', '0'), ('2', '0'), ('', '0'), ('Apr', '0'), ('18', '0'), ('', '0'), ('8', '0'), ('', '0'), ('total', '0'), ('111', '0'), ('', '0'), ('8', '0'), ('(', '0'), ('82', '0'), ('', '0'), ('7', '0'), ('', '0'), ('', '0'), ('Linseed', '0'), ('prev', '0'), ('34', '0'), ('', '0'), ('8', '0'), ('', '0'), ('Feb', '0'), ('32', '0'), ('', '0'), ('9', '0'), ('', '0'), ('Mar', '0'), ('6', '0'), ('', '0'), ('8', '0'), ('', '0'), ('Apr', '0'), ('6', '0'), ('', '0'), ('3', '0'), ('', '0'), ('total', '0'), ('88', '0'), ('', '0'), ('8', '0'), ('0'), ('(', '0'), ('87', '0'), ('', '0'), ('4', '0'), ('', '0'), ('', '0'), ('Soybean', '0'), ('prev', '0'), ('108', '0'), ('', '0'), ('9', '0'), ('', '0'), ('Feb', '0'), ('45', '0'), ('', '0'), ('1', '0'), ('', '0'), ('Mar', '0'), ('nil', '0'), ('', '0'), ('Apr', '0'), ('nil', '0'), ('', '0'), ('May', '0'), ('28', '0'), ('', '0'), ('', '0'), ('8', '0'), ('', '0'), ('total', '0'), ('166', '0'), ('', '0'), ('11', '0'), ('(', '0'), ('218', '0'), ('', '0'), ('5', '0'), ('', '0'), ('', '0'), ('Sunflowerseed', 'crop'), ('prev', '0'), ('48', '0'), ('', '0'), ('6', '0'), ('', '0'), ('Feb', '0'), ('61', '0'), ('', '0'), ('5', '0'), ('', '0'), ('Mar', '0'), ('25', '0'), ('', '0'), ('1', '0'), ('', '0'), ('Apr', '0'), ('14', '0'), ('', '0'), ('5', '0'), ('', '0'), ('total', '0'), ('149', '0'), ('', '0'), ('8', '0'), ('(', '0'), ('145', '0'), ('', '0'), ('3', '0'), ('', '0'), ('', '0'), ('Vegetable', '0'), ('oil', '0'), ('registrations', '0'), ('were', '0'), (':', '0'), ('Sunoil', '0'), ('prev', '0'), ('37', '0'), ('', '0'), ('4', '0'), ('', '0'), ('Feb', '0'), ('107', '0'), ('', '0'), ('3', '0'), ('', '0'), ('Mar', '0'), ('24', '0'), ('', '0'), ('5', '0'), ('', '0'), ('Apr', '0'), ('3', '0'), ('', '0'), ('2', '0'), ('', '0'), ('May', '0'), ('nil', '0'), ('', '0'), ('Jun', '0'), ('18', '0'), ('', '0'), ('8', '0'), ('', '0'), ('total', '0'), ('182', '0'), ('', '0'), ('4', '0'), ('(', '0'), ('117', '0'), ('', '0'), ('6', '0'), ('0'), ('', '0'), ('', '0'), ('Linoil', '0'), ('prev', '0'), ('15', '0'), ('', '0'), ('9', '0'), ('', '0'), ('Feb', '0'), ('23', '0'), ('', '0'), ('6', '0'), ('', '0'), ('Mar', '0'), ('28', '0'), ('', '0'), ('4', '0'), ('', '0'), ('Apr', '0'), ('2', '0'), ('', '0'), ('8', '0'), ('', '0'), ('total', '0'), ('61', '0'), ('', '0'), ('8', '0'), ('', '0'), ('(', '0'), ('76', '0'), ('', '0'), ('1', '0'), ('', '0'), ('', '0'), ('Soybean', '0'), ('oil', '0'), ('prev', '0'), ('3', '0'), ('', '0'), ('7', '0'), ('', '0'), ('Feb', '0'), ('21', '0'), ('', '0'), ('1', '0'), ('', '0'), ('Mar', '0'), ('nil', '0'), ('', '0'), ('Apr', '0'), ('2', '0'), ('', '0'), ('8', '0'), ('', '0'), ('May', '0'), ('9', '0'), ('', '0'), ('8', '0'), ('', '0'), ('Jun', '0'), ('13', '0'), ('', '0'), ('8', '0'), ('', '0'), ('', '0'), ('Jul', '0'), ('7', '0'), ('', '0'), ('8', '0'), ('', '0'), ('total', '0'), ('55', '0'), ('', '0'), ('8', '0'), ('(', '0'), ('33', '0'), ('', '0'), ('7', '0'), ('', '0'), ('', '0'), ('REUTER', '0')]

Named Entity Recognition-

Named Entity Recognition (NER) : {(Sunflowerseed', 'crop'), ('wheat', 'crop'), ('1986', 'year'), ('tonnes', 'unit')}

Measured Entity Recognition-

Measured Entity Detection: set()

For “reuters/training/6”, “1986/87” and “1985/86” are the years in the format (YYYY/YY) that are not listed in the year hence, lies under false positive.