UNIVERSITÄT WIEN
CSLEARN - EDUCATIONAL TECHNOLOGIES

Natural Language Processing

Exercise Sheet 7

Extracting Information from Text

Exercise 1

Extend the chunk grammar for the NP-chunker in Figure 2.2 to also match noun phrases containing plural head nouns. Test your grammar with the following sentences:

```
[("many", "JJ"), ("dogs", "NNS"), ("barked", "VBD"), ("at", "IN"),
  ("the", "DT"), ("cat", "NN")]
  => (S (NP many/JJ dogs/NNS) barked/VBD at/IN (NP the/DT cat/NN))

[("two", "CD"), ("dogs", "NNS"), ("barked", "VBD"), ("at", "IN"),
  ("the", "DT"), ("cat", "NN")]
  => (S (NP two/CD dogs/NNS) barked/VBD at/IN (NP the/DT cat/NN))

[("both", "DT"), ("new", "JJ"), ("dogs", "NNS"), ("barked", "VBD"),
  ("at", "IN"),  ("the", "DT"), ("cat", "NN")]
  => (S (NP both/DT new/JJ dogs/NNS) barked/VBD at/IN (NP the/DT cat/NN))
```

Exercise 2

Extend the grammar from Exercise 1 to also cover noun phrases that contain gerunds. Test your grammar with the following sentences and the sentences from Exercise 1:

```
[("many", "JJ"), ("dogs", "NNS"), ("barked", "VBD"), ("at", "IN"),
  ("the", "DT"), ("meowing", "VBG"), ("cat", "NN")]
(S
  (NP many/JJ dogs/NNS)
  barked/VBD
  at/IN
  (NP the/DT meowing/VBG cat/NN))

[("the", "DT"), ("man", "NN"), ("wants", "VBZ"), ("to", "TO"),
  ("become", "VB"), ("assistant", "NN"), ("managing", "VBG"),
  ("director", "NN")]
(S
  (NP the/DT man/NN)
  wants/VBZ
  to/TO
  become/VB
  (NP assistant/NN managing/VBG director/NN))
```

Exercise 3

Extend the grammar from Exercise 2 to also handle coordinated noun phrases. Test your grammar with the following sentences and the sentences from Exercise 1 and 2:

```
[("the", "DT"), ("man", "NN"), ("wants", "VBZ"), ("to", "TO"),
  ("leave", "VB"),  ("in", "IN"), ("July", "NNP"),
  ("or", "CC"), ("August", "NNP")]
(S
  (NP the/DT man/NN)
  wants/VBZ
  to/TO
  leave/VB
  in/IN
  (NP July/NNP or/CC August/NNP))

[("Donald", "NNP"), ("fired", "VBD"), ("all", "PDT"),
  ("your", "PRP$"), ("managers", "NNS"), ("and", "CC"),
  ("supervisors", "NNS")]
(S
  (NP Donald/NNP)
  fired/VBD
  (NP all/PDT your/PRP$ managers/NNS and/CC supervisors/NNS))

[("company", "NN"), ("personnel", "NN"), ("policy", "NN"), ("has", "VBZ"),
  ("always", "RB"), ("been", "VBN"), ("the", "DT"), ("law", "NN"),
  ("that", "WDT"), ("rules", "VBZ"), ("company", "NN"),  ("courts", "NN"),
  ("and", "CC"), ("adjudicators", "NNS")]
(S
  (NP company/NN personnel/NN policy/NN)
  has/VBZ
  always/RB
  been/VBN
  (NP the/DT law/NN)
  that/WDT
  rules/VBZ
  (NP company/NN courts/NN and/CC adjudicators/NNS))
```

Exercise 4

Extend the chunk grammar from Exercise 1 to a multi-stage chunk grammar for a cascaded chunker that produces the following output for the three test sentences:

```
(S
  (NP many/JJ dogs/NNS)
  (VP barked/VBD (PP at/IN (NP the/DT cat/NN))))
```

```
(S
  (NP two/CD dogs/NNS)
  (VP barked/VBD (PP at/IN (NP the/DT cat/NN))))
(S
  (NP both/DT new/JJ dogs/NNS)
  (VP barked/VBD (PP at/IN (NP the/DT cat/NN))))
```

Exercise 5

Write a Definite Clause Grammar (DCG) in SWI-Prolog, which corresponds to the chunk grammar in Exercise 4 and produces the following output:

```
[np([many/'JJ',dogs/'NNS']),vp([barked/'VBD',
  pp([at/'IN',np([the/'DT',cat/'NN'])])])]

[np([two/'CD',dogs/'NNS']),vp([barked/'VBD',
  pp([at/'IN',np([the/'DT',cat/'NN'])])])]

[np([both/'DT',new/'JJ',dogs/'NNS']),vp([barked/'VBD',
  pp([at/'IN',np([the/'DT',cat/'NN'])])])]
```

Exercise 6

Write a predicate `print_tree(Tree)` in SWI-Prolog that produces the following output for the three sentences from Exercise 5:

```
 (S
    (NP many/JJ dogs/NNS)
    (VP barked/VBD
        (PP at/IN
            (NP the/DT cat/NN))))

(S
    (NP two/CD dogs/NNS)
    (VP barked/VBD
        (PP at/IN
            (NP the/DT cat/NN))))

(S
    (NP both/DT new/JJ dogs/NNS)
    (VP barked/VBD
        (PP at/IN
            (NP the/DT cat/NN))))
```

The following predicates may be useful for the implementation:

a) `Term=..List`: `List` is a list whose head is the functor of `Term` and the remaining arguments are the arguments of the term, e.g. `?- T =..[a, b, c].  T=a(b,c).`;

b) `upcase_atom(Atom, UpperCase)`: converts an atom to uppercase;

c) `tab(Amount)`: writes `Amount` spaces on the current output stream.

Exercise 7

Extend the multi-stage chunk grammar from Exercise 4 to also cover the test sentences from Exercise 2:

```
(S
  (NP many/JJ dogs/NNS)
  (VP barked/VBD (PP at/IN (NP the/DT meowing/VBG cat/NN))))

(S
  (NP the/DT man/NN)
  (VP
    wants/VBZ
    (INFCL
      to/TO
      (VP become/VB (NP assistant/NN managing/VBG director/NN)))))
```

Exercise 8

Extend the Definite Clause Grammar from Exercise 5 to also cover the test sentences from Exercise 2:

```
(S
    (NP many/JJ dogs/NNS)
    (VP barked/VBD
        (PP at/IN
            (NP the/DT meowing/VBG cat/NN))))

(S
    (NP the/DT man/NN)
    (VP wants/VBZ
        (INFCL to/TO
            (VP become/VB
                (NP assistant/NN managing/VBG director/NN)))))
```

Exercise 9

Extend the multi-stage chunk grammar from Exercise 7 to also cover the test sentences from Exercise 3:

```
(S
  (NP the/DT man/NN)
  (VP
    wants/VBZ
    (INFCL
      to/TO
      (VP leave/VB (PP in/IN (NP July/NNP or/CC August/NNP))))))
(S
  (NP Donald/NNP)
  (VP
    fired/VBD
    (NP all/PDT your/PRP$ managers/NNS and/CC supervisors/NNS)))

(S
  (NP company/NN personnel/NN policy/NN)
  (VP
    has/VBZ
    always/RB
    been/VBN
    (NPRC
      (NP the/DT law/NN)
      (RELCL
        that/WDT
        (VP
          rules/VBZ
          (NP company/NN courts/NN and/CC adjudicators/NNS))))))
```

Exercise 10

Extend the Definite Clause Grammar from Exercise 8 to also cover the test sentences from Exercise 3:

```
(S
    (NP the/DT man/NN)
    (VP wants/VBZ
        (INFCL to/TO
            (VP leave/VB
                (PP in/IN
                    (NP July/NNP or/CC August/NNP))))))
```

```
(S
    (NP Donald/NNP)
    (VP fired/VBD
        (NP all/PDT your/PRP$ managers/NNS and/CC supervisors/NNS)))

(S
    (NP company/NN personnel/NN policy/NN)
    (VP has/VBZ always/RB been/VBN
        (NPRC
            (NP the/DT law/NN)
            (RELCL that/WDT
                (VP rules/VBZ
                    (NP company/NN courts/NN and/CC adjudicators/NNS))))))
```

Exercise 11

Select the VP chunks from the "train" portion of the CoNLL 2000 Chunking Data Corpus. Create a list of all the tag sequences that occur with each instance of this chunk type. Use a `FreqDist` to display the 100 most common tag sequences.