School of Computing and Information Systems
The University of Melbourne
COMP90042
WEB SEARCH AND TEXT ANALYSIS (Semester 1, 2017)

Workshop exercises: Week 4

Discussion

1. What is **chart parsing**? Why is it important?
2. Consider the following simple **context–free grammar**:

```
S -> NP VP
VP -> V NP | V NP PP
PP -> P NP
V -> "saw" | "walked"
NP -> "John" | "Bob" | Det N | Det N PP
Det -> "a" | "an" | "the" | "my"
N -> "man" | "cat" | "telescope" | "park"
P -> "on" | "by" | "with"
```

   (a) What changes need to be made to the grammar to make it suitable for **CYK parsing**?
   (b) Using the CYK strategy and the above grammar in CNF, parse the following sentences:
      i. "a man saw John"
      ii. "an park by Bob walked an park with Bob"
      iii. "park by the cat with my telescope"

3. What is a **probabilistic grammar** and what problem does it attempt to solve?

Programming

1. Using the framework from the `WSTA_N5_context-free_grammars` iPython notebook, input the grammar and parse the sentences given in the Discussion. Are the results what you expected?
2. How many parses are there for the sentence "revenue increased last quarter", based on the Penn Treebank corpus? Why are there so many?
3. Work through the probabilistic CYK parser given in the `WSTA_N6_probabilistic_parsing` iPython notebook. Ensure that you understand how the tree can be built from the chart.

Catch-up

- What is a **grammar**? What is **parsing**?
- Revise the syntax for rules in a **context–free grammar**. In particular, familiarise yourself with the terms **terminal**, **non-terminal**, **productions**, **start symbol**, **syntax tree**.
- What is a **constituent**? What is the significance of the following: **Noun Phrase**, **Verb Phrase**, **Prepositional Phrase**, **Adjective Phrase**, **Adverbial Phrase**, **Subordinate Clause**?
- What is the difference between **top-down** and **bottom-up** parsing?
- How can a **prior** probability be estimated from a collection of data, using a **maximum likelihood estimate** approach? What about a **posterior** probability?
- Why are we often concerned by a model where some events have a probability equal to 0?

Get ahead

- Parse the sentences from the Discussion section using the Earley strategy.
- Work through the probabilistic Earley algorithm example given in the NLTK PCFG demo.