Discussion

Pre-workshop

• What is a “grammar”?
• What is “parsing”? What is the difference between “top-down” and “bottom-up” parsing?

Workshop

1. Consider the following English sentence:

   ??I saw the field the gate the horse the girl the boy liked rode jumped enclosed.

   (a) What is the significance of the ?? marker at the start of the sentence?
   (b) What property of language is being demonstrated here?

2. Consider the following simple context-free grammar:

   \[ 
   \begin{align*} 
   S & \rightarrow NP \ VP \\
   VP & \rightarrow V \ NP \mid V \ NP \ PP \\
   PP & \rightarrow P \ NP \\
   V & \rightarrow \text{"saw"} \mid \text{"walked"} \\
   NP & \rightarrow \text{"John"} \mid \text{"Bob"} \mid \text{Det} \ N \mid \text{Det} \ N \ PP \\
   \text{Det} & \rightarrow \text{"a"} \mid \text{"an"} \mid \text{"the"} \mid \text{"my"} \\
   N & \rightarrow \text{"man"} \mid \text{"cat"} \mid \text{"telescope"} \mid \text{"park"} \\
   P & \rightarrow \text{"on"} \mid \text{"by"} \mid \text{"with"} 
   \end{align*} \]

   (a) Parse the following sentences:
   i. “a man saw John”
   ii. “an park by Bob walked an park with Bob with an park”
   iii. “park by the cat with my telescope”
   (b) Are there any problems with the grammar above? How could you fix them?

3. What is “chart parsing”? What problem is it trying to solve?

   (a) Identify the edges of the “chart” created when parsing the “terminals” of the sentences above.
   (b) Identify the edges of the “chart” created when parsing the “non-terminals”, for the sentences above. What is a “pre-terminal”, and why is it often treated differently?
   (c) Discuss the role of the “dotted” notation.

Post-workshop

• Revise the basic rules of chart parsing. Create one or more sentences based on the grammar above, and parse them.
Programming

Pre-workshop


Workshop

1. Using NLTK (section 3.1), parse the sentences above using the grammar above.

2. Make some alterations to the grammar and try to parse those sentences again. Confirm that the behaviour of the parser is as expected.

Post-workshop

- Write some more sentences and parse them. Alter your grammar as necessary.