

Homework #3

Due: 17 November 2011, by 13:00h.

Exercise 1 The TUNA Reference Corpus is a freely available corpus of referring expressions to objects in visual domains: <http://inf.abdn.ac.uk/research/tuna/corpus/>

The corpus was collected through controlled experiments run online. Subjects were shown scenes depicting a set of objects, one of which was the target. They were asked to type in a description of the target for an “imaginary” addressee. The corpus has been used, amongst other things, to evaluate different algorithms for the generation of referring expressions against a human gold standard.

In this directory <http://staff.science.uva.nl/~raquel/teaching/rid/tunaexamples/> you will find two HTML files, each showing a sample TUNA scene with a description generated by one of the human subjects. Each scene and description are annotated in the corresponding XML file. To examine the annotation, open the XML files with a text editor (your browser may not display the XML code, which you want to see). The XML files should be self-explanatory, but in case you need further details regarding the XML annotation format, you can find them in the technical report by Gatt, van der Sluis & van Deemter (2008), which is available from the website of the corpus.

Please answer the following questions and send me a PDF by email (see deadline above):

- 1.1 Are these two sample descriptions minimal? Explain your answers in terms of the definitions given in class.
- 1.2 Which list(s) of preferred attributes would allow the Incremental Algorithm to generate these descriptions? Again, explain your answers in terms of the definitions given in class.