8 Natural Language Processing (SHT)

(a) How many senses of “plant” can you identify in the snippets from the 4 documents given above? Describe each sense by giving a hypernym. [2 marks]

(b) Describe how Yarowsky’s algorithm for word sense disambiguation would process the example texts. Illustrate each stage of the algorithm with an example. [6 marks]

(c) Under adverse conditions, the algorithm in part (b) can rapidly diverge from a good solution. What are these conditions? Illustrate using the examples above. [4 marks]

(d) How could you apply the Naive Bayes machine learning algorithm, discussed in the lectures in combination with several tasks, as a method for learning the word senses for an ambiguous word? Give the formula and explain how the necessary parameters can be trained. [8 marks]