Master Degree in Computer Engineering

Natural Language Processing
Final Exam

September 20th, 2023

1. [**2 points**]   Introduce Herdan/Heaps law and discuss its relevance for natural language processing.

2. [**6 points**]   Consider the following term-context matrix, providing co-occurrences for the target words oven, algorithm, door and kitchen, against the contexts words school, computer, spoon and house.

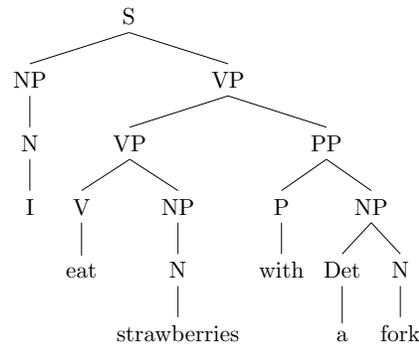|	school	computer	spoon	house
oven	3	0	11	9
algorithm	25	105	0	2
door	5	0	0	87
kitchen	1	0	78	117

   For all entries in the main diagonal, indicate how to compute the positive pointwise mutual information. Use fractions and logarithms in your answers **without** computing these operators. For the purpose of this exercise, assume that no other word/context pairs matter, and assume that each context word appears in the context of only one occurrence of a target word, and the other way around.

3. [**5 points**]   With reference to the contextualized language models, answer the following questions.

   (a) Introduce the basic architecture of BERT.
   (b) Define and motivate the two training methodologies of masked language modeling and next sentence prediction.

4. [**2 points**]   Describe the task of named entity recognition (NER), and define the BIO tagging approach, discussing a simple example.

5. **[5 points]** With respect to syntactic representations of natural language, answer the following questions.

   (a) Describe the two steps algorithm presented in our lectures for the conversion from phrase structure tree to dependency tree.

   (b) Apply the algorithm to the following phrase structure tree, providing and discussing also the representation obtained at the intermediate step.

```
                              S
                  ┌───────────┴───────────┐
                 NP                       VP
                  │                ┌───────┴───────┐
                  N               VP               PP
                  │           ┌────┴────┐      ┌────┴────┐
                  I          V         NP      P        NP
                             │          │      │     ┌───┴───┐
                            eat         N     with   Det     N
                                        │             │      │
                                   strawberries        a    fork
```

6. **[6 points]** Introduce the neural model known as Stanford attentive reader for the task of machine reading. Define the main representation and equations used by this model.

7. **[5 points]** With reference to digital assistants, answer the following questions.

   (a) Introduce the so-called frame-based architecture and the three main tasks of domain classification, intent determination, and slot filling.

   (b) Specify the technique of semantic grammars, providing some simple examples.

8. **[2 points]** With reference to linguistic theory, outline the distinction between the field of study of general semantics and the field of study of pragmatics.