1. In Problem 3, you did an error analysis on the task of recognizing text spans mentioning named entities. Pick one type of error that you observed. How could you improve the model's performance on this type of error? What resources (such as domain knowledge, data, compute, ...) would you need to implement this improvement?

   The model faced difficulty to identify named entities correctly and identified dates, numbers and abbreviations wrongly. Model predicted numbers and dates as named entities. Also, it was confused with acronyms and struggle to understand the context of some phrases related to time. One of the observations is "Sentence: Germany imported 47,600 sheep from Britain last year , nearly half of total imports .   FP: 47,600, FP: nearly half, FP: last year ". To improve the performance, a filter was implemented which contains a list of labels (MONEY, PERCENT, TIME, DATE, CARDINAL, ORDINAL, QUANTITY). Any identified entities falling into these categories are excluded from the final predictions. This approach reduced false positives.

2. What does the word "context" refer to in the context of Problem 6? How does this help to disambiguate between different entities? Suggest another type of context that you could use for disambiguation.

   The word "context" refers to the surrounding textual information. For the data given, a context is defined as the 5 tokens to the left and the 5 tokens to the right of the mention, used for disambiguating between different entities. This helps to figure out which thing we're talking about by looking at the words around it. It helps predict the right thing based on how it's used in a sentence. Another type of context could be considering syntactic relationships which involve analyzing the grammatical structure of the sentence.

3. Why One type of entity mentions that we did not cover explicitly in this lab are pronouns. As an example, consider the following sentence pair:
   Ruth Bader Ginsburg was an American jurist. She served as an associate justice of the Supreme Court from 1993 until her death in 2020.
   What facts would you want to extract from this sentence pair? How do pronouns make fact extraction hard?

   From the sentences, we want to know the information about Ruth Bader Ginsburg, her profession as a justice, and when did she serve in the Supreme Court. Pronouns like "She" make it tricky because we need to figure out who or what they're talking about. If we misunderstand, we might get the facts wrong.