

RQ 4.1:

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 tend to recognize numbers as an entity. We can just simply remove all numerical strings since in most cases they are all FP as we can see from the df_kb knowledge base as a prior knowledge. So it's safe to do that.
```

RQ 4.2:

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 'context' here means consecutive nearby words in order, in the words sequence we have regarding the corresponding mention(and the corresponding entity). Here we use at most 5 words per direction for the mention as context. And by training classifier from the context we can learn a model help us predict given a new context of a mention, which most possible entity will it be.  
we can also use nearby #n recognized entitys instead of full texts as context, or more advanced information we extracted from nearby texts such as topic.
```

RQ 4.3:

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?

```
we want to extract that "Ruth Bader Ginsburg served as an associate justice".  
But which is pronouns' target sometimes confusing even for human. It might be far away in the text from the nouns they are refering, and there might have multiple possible nouns for pronouns to refer.
```