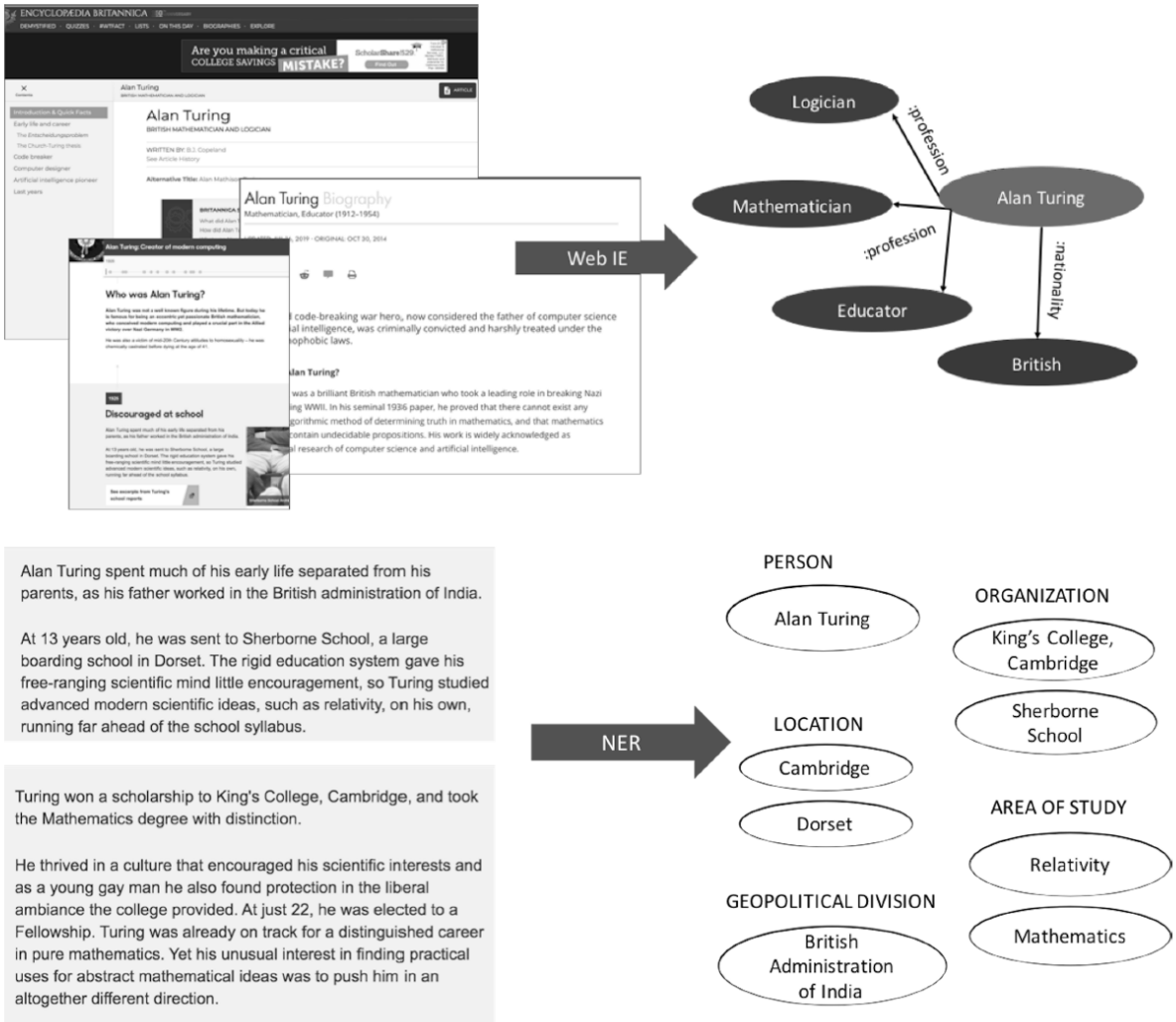
Review: Named Entity Recognition (NER)

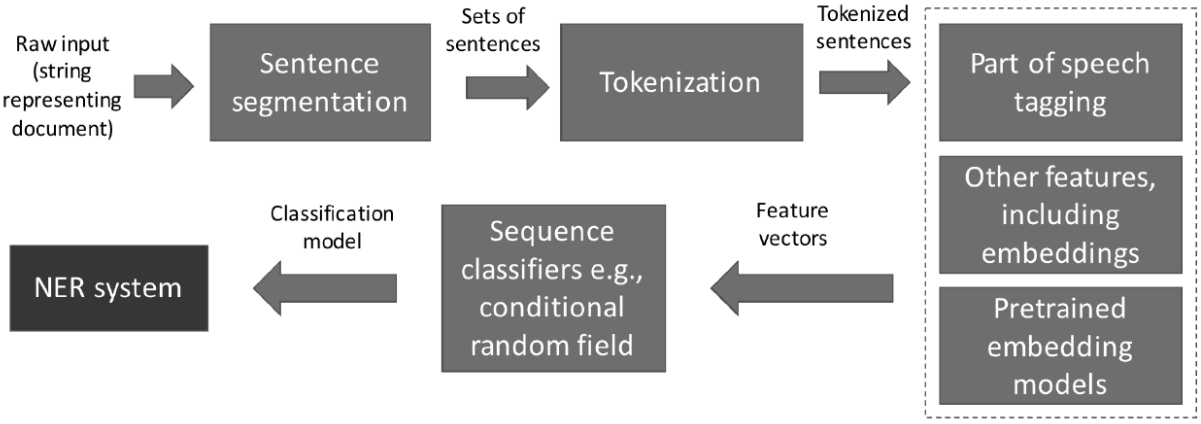
Let's start by reviewing Web IE vs. NER through an example:



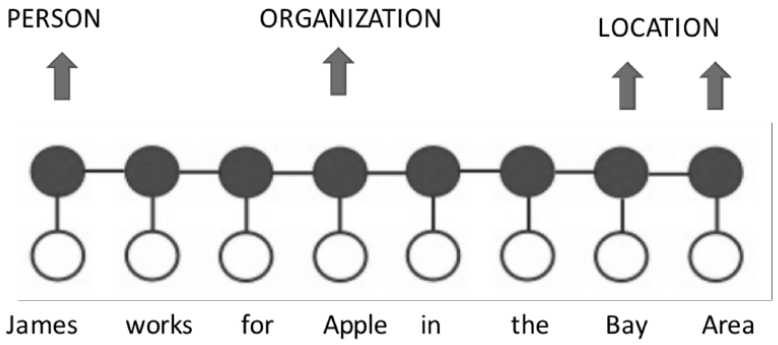
What are some challenges that Web IE faces that NER does not necessarily face? And vice versa?

Why is IE hard? List some specific challenges and provide examples to illustrate your case more clearly.

Practical architecture for NER (make sure you understand all the components below at least at the qualitative level):



Recall conditional random fields (CRFs):

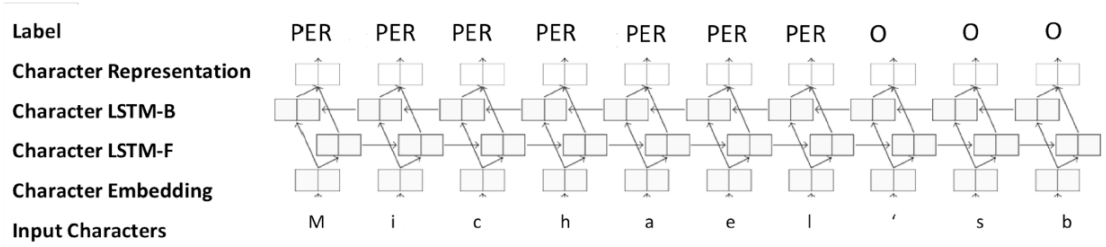


In the final, you would not be expected to know many formulae but it's not a bad idea to remember them, as they can prove useful (make sure you know what the symbols mean; if not, it's a good idea to review the lecture material):

$$p(s|x;w) = \frac{e^{w.\Phi(x,s)}}{\sum_{s'} e^{w.\Phi(x,s')}}.$$

$$L(w) = \sum_{i=1}^{n} logp(s^i|x^i;w) - \frac{\lambda_2}{2}\|w\|_2^2 - \lambda_1\|w\|_1.$$

Make sure you understand the components in the diagram below at a qualitative level:



A character-based RNN architecture as applied to NER for an input sentence such as "Michael's birthday is coming."

Domain-specific IE: what is it and why is it important? Can you give some illustrative use cases?

IE quality metrics

**make sure you know what the symbols below mean. If not, then you should review lecture material again*

$$Precision = \frac{\#correct}{\#correct + \#incorrect}.$$

$$Recall = \frac{\#correct}{\#total}.$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}.$$

$$SER = \frac{\#incorrect + \#missing}{\#total}.$$

Advanced topics: Relation extraction and event extraction

What are some ontologies for relation extraction?
Why are RE and EE harder than NER?