

Homework 6

58119125 蒋卓洋

1. Please compute the Levenshtein distance between two strings “Saturday” and “Sunday”.

Solution:

$$\begin{aligned}LD &= \text{lev}(\text{“Saturday”}, \text{“Sunday”}) \\&= \text{lev}(\text{“aturday”}, \text{“unday”}) \\&= 2 + \text{lev}(\text{“urday”}, \text{“unday”}) \\&= 2 + \text{lev}(\text{“rday”}, \text{“nday”}) \\&= 3 + \text{lev}(\text{“day”}, \text{“day”}) \\&= 3\end{aligned}$$

2. Please compute the Jaccard similarity between “French” and “France” based on character-level bi-grams.

Solution:

$$\begin{aligned}\text{Bi-grams_Segmentation}(\text{“French”}) &= \{\text{Fr, re, en, nc, ch}\} \\ \text{Bi-grams_Segmentation}(\text{“Frence”}) &= \{\text{Fr, re, en, nc, ce}\} \\ J(\text{“French”}, \text{“France”}) &= \frac{|\{\text{Fr, re, en, nc}\}|}{|\{\text{Fr, re, en, nc, ch, ce}\}|} = \frac{4}{6} = \frac{2}{3}\end{aligned}$$

3. Please tell the difference between local disambiguation and global disambiguation in entity linking.

Solution:

- (1) Local disambiguation just do disambiguation according to the contextual information of the given mention and the target entity without considering the effect of other referent entities and other string mentions.
- (2) Global disambiguation will consider the effect of all referent entities and string mentions, leveraging the semantic associations between global candidate entities to jointly disambiguate all entities for the mentions in a table.

4. Why do we need candidate generation in entity linking?

Solution:

Because for a string mention itself, it may represent different entities in different context, so we need to find some possible entities in the KB .