Assignment 10

Due on 2019-10-30, 23:59 UTC

The data scientists in the assignment have access to the dataset.

1. The times ‘Today’s’ sections in Wikipedia (1000 times), among which ‘SWAT’ data is listed at the end of the list. The times ‘C’ is a set of 1000 times it is used in the context of a programming language. Which one of the times ‘C’ and ‘Today’s’ has higher frequency? (Choose one)

2. $\frac{1}{x} + \frac{1}{y} = \frac{1}{z}$

3. $\frac{1}{3} + \frac{1}{4} = \frac{1}{5}$

4. Not enough information available for calculation of Keylogger.

5. You are trying to disambiguate between three different forms of the word “NEP”.
   - Version 1: English
   - Version 2: Non-Linear Programming
   - Version 3: Non-Linear Programming
   - Which is the true form for the words “NEP”? (Choose one)

6. Background between the context of a key word/phrases and possible links are measured using:
   1. Context words and phrases
   2. Context (heuristics) and phrases
   3. Context (heuristics) and context (dependence on the content of the keyword)
   4. Context (heuristics) and context (dependence on a sentence and all filled parts appearing in the content of the key word/phrases)
   5. Context of the target link

7. Work of the following sequence of steps for entity linking is correct:
   1. Disambiguation → Link Generation → Relation Detection
   2. Link Generation → Relation Detection → Disambiguation
   3. Relation Detection → Link Generation → Disambiguation
   4. Relation Detection → Disambiguation → Link Generation

8. Distinct supervised relation extraction mechanisms have an advantage of:
   1. Supervised approaches by using reliable head-landmark knowledge
   2. Unsupervised approaches by utilizing undefined amounts of text data
   3. Both a and b
   4. None of the above

9. Relevant features for a supervised model for predicting the topics to be judged (true)
   1. Disambiguation Confidence
   2. Relations
   3. Topic match between the text entities/logic
   4. All of the above

10. Work of the following patterns can be used for obtaining remaining unparseable / incomplete extensions
    1. X including Y
    2. X such as Y
    3. X or Y
    4. X and Y

11. In Entity Supervision approach for learning a model for relation extraction, how do we obtain negative samples by training the classifier?
    1. Expert annotated negative samples in test corpus
    2. Negative samples are created using unmarked entity pairs in knowledge base
    3. Negative samples are acquired using supervision
    4. None of the above