

Natural Language Processing

Assignment 10

Type of Question: MCQ

Number of Questions: 10 Total Marks: $(10 \times 1) = 10$

1. Different phases of entity linking are -

- A) Candidate Selection -> Reference Disambiguation
- B) Reference Disambiguation -> Candidate Selection -> Mention Identify
- C) Mention Identify -> Candidate Selection -> Reference Disambiguation
- D) All of the above

[Marks 1]

Answer: C

Solution: Theory. Slide 7, Lecture 1, Week 10 (Entity Linking).

2. The text span s="Sea" occurs in 600 different Wikipedia articles.

c1	223
c2	161
c3	18
c4	11
No Link	187

Calculate the keyphraseness of "Sea".

[Marks 1]

Answer:

- A) 0.232
- B) 0.886
- C) 0.688
- D) 0.976

Solution: C) $CF(s_i) / CF(s) = 223 + 161 + 18 + 11 / 600 = 413 / 600 = 0.688$

3. What is the commonness of (s, c2) in the above question?

[Marks 1]

Answer:

- A) 0.765
- B) 0.389
- C) 0.145
- D) 0.910

Solution: B) $161/(223+161+18+11) = 161/413 = 0.389$

4. Relevant feature/s for a supervised model for predicting the topics to be linked is/are:

- A) Disambiguation Confidence
- B) Relatedness
- C) Link Probability
- D) All of the above

[Marks 1]

Answer: D

Solution: Theory. Entity Linking Lecure II [Lecture Video 47]

5. Which of the following problem exists in bootstrapping technique for Information extraction are:

- A) Sensitiveness towards the seed set
- B) High precision
- C) Less manual intervention
- D) All of the above

[Marks 1]

Answer: A

Solution: Theory

6. Which of the following is an advantage of unsupervised relation extraction:

- A) Can work efficiently with small amount of hand-labelled data
- B) Not easily generalizable to different relations
- C) Need no training data.
- D) Always perform better than supervised techniques.

[Marks 1]

Answer: C

Solution: Theory

7. Which of the following is not a Hearst's Lexico Syntactic Patterns for automatic acquisition of hyponyms -

- A) X or other Y
- B) X and other Y
- C) Y including X
- D) X but not Y

[Marks 1]

Answer: D

Solution: Theory. Lecture Video 48 : Information Extraction - Introduction, Week 10, Lecture 3 - Slide 15/18.

8. Advantage of Distant supervision over bootstrapping method

- A) Need more data
- B) Less human effort
- C) Can handle noisy data better
- D) No Advantage

[Marks 1]

Answer: C

Solution: Theory

9. Consider a dataset with a very low number of relations - all of which are very important. For a relation extraction task on that dataset, which of the following is the most useful metric

- A) Precision
- B) Recall
- C) Accuracy
- D) F1-Score

[Marks 1]

Answer: B

Solution: Theory

10. What is KeyPhraseness (wikipedia)?

- A) Number of articles that mention a key phrase divided by the number of wikipedia articles containing it.
- B) Number of Wikipedia articles that use it as an anchor, divided by the number of articles that mention it at all.
- C) Number of articles that mention a key phrase times by the number of wikipedia articles containing it.
- D) Number of Wikipedia articles containing the key phrases times by number of articles mentioning it.

[Marks 1]

Answer: B

Solution: Theory.
