

Student ID _____ :

Name-Surname _____ :

Natural Language Processing Final exam (2019-2020 Fall)

(Please use free space for draft and fit your answer to boxes.)

1. Fill in the blanks with the most appropriate terms. Each question is five points (5P).

- a. _____ is a semantic relation that used when two words exactly or nearly the same.
- b. _____, which is mostly used in spell checking, is a metric to measure the difference between two words.
- c. _____, which is a feature extraction method from a text, calculates the number of repetitions for each word depending on the number of documents.
- d. _____ similarity deals with positions of characters.
- e. WordNet is the most used tool to solve semantic _____ problem.
- f. The word vectors computed by co-occurrence statistics can be used in _____ level semantic similarity.
- g. _____ is used for checks the quality of chatbots.
- h. _____ based question answering techniques extract passages directly from documents, guided by the text of the question.
- i. In _____ machine translation, we need firstly a target language model and translation model. Then, we can calculate probabilities.
- j. The first Neural based translation system is developed by _____.

2. Giving your comments, explain your solution. Each question is twenty-five points (25P).

a. Let Q be a short question of only one sentence. In a very large corpus, how can we find the semantically most relevant text to this question?

b. Let Q be a machine translation project to do. In this project, the most important problems are related to ambiguities. For example, when user enters "The fisherman went to the bank", although the word "bank" means "coast" here, the designed software can translate it as a word that means financial institution. How can you solve this problem?