

9 Natural Language Processing (ES)

The goal of automatic summarisation is to produce a short version of a text that contains the most important or relevant information. In multi-document summarisation, we need to aggregate content from multiple documents into one cohesive summary.

(a) Describe two ways in which multi-document summarisation is more challenging than single-document summarisation. [4 marks]

(b) In query-focused multi-document summarisation, sentences can be selected for the summary based on maximal marginal relevance (MMR). Give the formula for MMR and explain the intuition behind it. [5 marks]

(c) The two sentences in each sentence pair below are linked by a particular rhetorical relation. Which rhetorical relation does each sentence pair exhibit?

(i) *The use of diesel in transport has come under increasing scrutiny in recent years. According to WHO, around three million deaths every year are linked to exposure to outdoor air pollution.*

(ii) *Nitrogen oxides can help form ground level ozone. This can exacerbate breathing difficulties.*

(iii) *Paris has already taken a series of steps to cut the impact of diesel cars and trucks. Vehicles registered before 1997 have already been banned from entering the city.*

[1 mark each]

(d) Briefly discuss how each of the following NLP techniques can be used in extractive summarisation.

(i) morphological processing;

(ii) syntactic parsing;

(iii) lexical and distributional semantics;

(iv) discourse parsing, i.e. identification of rhetorical relations.

[2 marks each]