

PGT Exam Performance Feedback

2016/2017 Semester 2

COMP61332 Text Mining

John McNaught
Riza Batista-Navarro

Comments Please see the attached report.

Comments on the examination

1. General comments

These comments are made in relation to unapproved scores.

28 candidates sat the examination. The average mark was 59.82%. The top mark was 80%, the bottom mark was 31.67%. 6 candidates scored under 50%. 9 marks were in the distinction range and 6 in the merit range. 23 candidates answered question 1 (av. ~64%); 27 candidates answered question 2 (av. ~62.5%); 21 answered question 3 (av. ~54%); 10 answered question 4 (av. 51.5%); and 5 answered question 5 (av. 55%). In several cases, parts were not attempted, which would clearly lead to depressed scores. Also, there were several vague, general, over-brief answers that did not persuade that the candidates really understood points at issue. Marks were lost by some, due to not answering all (especially discussion) parts of questions.

2. Specific comments (not all question parts merited comment)

Question 1

a) This is a bookwork question yet many did not get full marks mostly due to missing the categorical and garden path types of ambiguity.

b) i) Majority answered this question satisfactorily. The few candidates who got it wrong were under the impression that the non-noun phrase (“outside”) tokens should not be included.

b) ii) In general, this question was not addressed well. Many stated that performance depends on the data but only very few elaborated on their answer nor mentioned the possibility of sparsity, e.g., that the tokens that BILOU specifies better could be sparse in the data set thus the consequence that BILOU is any better than BIO.

d) i) Many students answered by discussing the advantages of text mining architectures rather than what their characteristics are, in which case they were given full marks if the discussion was thorough.

Question 2

a) This was generally not well tackled. Some gave no discussion at all or did not indicate what issues might arise. Others confused tokenization and NER (especially of multi-token entities).

b) i) Some did not consider the role of templates, or the role of iteration.

ii) Several answers failed to take account of the difference between training a tagger and applying a tagger. The question concerned applying a tagger: “used to tag a new, unseen text”. One does not do training at the same time as tagging, as these answers tried to convince me. It is important to read a question to understand what is being asked.

c) ii) was answered poorly by several as they drew a tree or several subtrees with no further comment, i.e., no explanation of steps according to the naïve top-down depth-first algorithm. Those that referred to the grammar and explained how a naïve top down depth-first parse would deal with backtracking did well. Some candidates attempted to explain how the sentences "winter storms strikes" or "winter storms delay trains" would be parsed, but *these were not the sentence of this question part*. It pays to read the question carefully.

e) Some candidates generated orphans (a standalone link between two words, but nothing else linking in to the overall graph). Others had wrong direction for a relation (this was treated leniently). Some had wrong labels or wrong dependencies, or left some words entirely unconnected. Consistent use of the wrong label or dependency was not multiply penalized.

Question 3

a) There were many underdeveloped answers to this question. Majority provided a discussion around contextual features but giving very limited examples (e.g., some mentioning only nouns (e.g., "flight", "match")). Many did not provide any comparison (e.g., giving a discussion only of the case where "Etihad" refers to the stadium but not saying anything about the case where it refers to the organisation).

b) i) This question was generally answered well, although a common mistake was to leave out the DIE event. Majority also specified the same entity ("Ian Stewart") for the DIE event's Agent argument and the ACCUSATION event's Defendant argument. Full marks were given if there was no redundancy, i.e., it was specified only in the latter.

b) ii) This question was not addressed well in general. Many provided context words (e.g., words within the neighbourhood of "charge") rather than syntactic relations, on which they have had laboratory exercises and coursework.

e) This question was not answered well in general. Majority seem to not have understood how facets/categories can help narrow down results in the context of search. However, some were able to link faceted search with categories derived from named entity recognition.

Question 4

a) Several candidates did not refer to WN relationships beyond synsets, or discuss disadvantages.

b) ii) This was generally not well tackled, with only partially correct results in the main.

c) i) Some gave only a very few features, for the marks available.

ii) Some gave very general answers that would be true of supervised machine learning in general, so did not address the particular issues in relation to WSD.

Question 5

b) This was not answered by some.

c) There were some over-brief (given the marks available) and general answers to this discussion-type question, and other very good ones that showed very good awareness.