Sandra Sampaio
Jock McNaught
Goran Nenadic

Comments

Question 1

Comments relate to unconfirmed results, and only to noteworthy question parts.

57 Candidates answered this question. The highest mark was 29/30, the lowest 9/30. Several question parts were generally the subject of over-answering and verbosity, however there were numerous other answers that were underdeveloped and general in nature.

Q1a) was generally well answered, although some wrote much more than necessary for the marks available. Some apparently misunderstood the question and wrote about document processing aspects, and did not address classification of needs, despite there having been a workshop on this.

Q1c iii) had mainly good answers although a few candidates did not relate their answer to that for 1c ii) as instructed.

Q1d) was generally well answered, although there were several underdeveloped answers for the marks available. However, several candidates confused a language recogniser with a tokeniser: a language recogniser does what it says, it recognises what language a text is written in. Several thought that the Porter stemmer could handle languages other than English. Several did not address the positional inverted index aspect.

Q1e i) was not well answered by many, who simply reproduced the IDF formula and lengthily explained it, but did not give an answer to the question. This is indicative of memorising as opposed to being knowledgeable, and is moreover a waste of time for a low-value part.

Q1e iii) again saw answers simply reproducing the tf-idf formula without answering the question. Lab work in particular should have provided a clear answer here.

Q1f ii) saw some very inventive but wrong answers to what was essentially a simple question.

Q1f iii) was well answered, but a few did not interpret the cosine value properly.

Q1g) was also well answered by many, although some did not attempt it. In several cases, candidates failed to arrive at correct values after 2 iterations, or managed to get only 1 of the 2 values correct, however I was able to award partial marks for some.

Q1i) EITHER: attracted some very good answers that demonstrated deep learning and moreover the ability to bring in contributions from other parts of the unit. Several answers remained very general/vague and did not show much knowledge of advanced techniques related to the unit's topics. Many answers were underdeveloped for the marks available: I hypothesise that this was due to one of two things, or indeed both: i) a lack of reading around on one's own initiative, of investigating recommended material (over-reliance on slides); ii) inability to produce an argument of the type expected from a final year student, demonstrating critical analysis and awareness of key issues.

Q1i) OR: was not well answered, with generally poor awareness being shown about the crucial role of context in helping to identify named entities. For example, several claimed that a sequence of initially capitalised tokens would indicate a PERSON name, which would mean e.g., "Clever Do You" being taken as a PERSON and "Abbou de Souffle" being missed. Some thought that 16th was a number, others thought that a DATE would be recognised by a sequence of day month year without specifying how these components would be recognised.

On several occasions, I could not decipher handwriting due to illegible words.

In answers, several students referred to topics of course workshops and brought in their experiences from these and the lab, all of which was very welcome. Some made very good references to aspects/applications beyond the immediate concerns of the course, equally welcome.

Question 2

Unfortunately, the majority of the marks for this question were in the average zone, and this was largely due to the poor quality of the provided answers to the various items. It has become quite obvious that the vast majority of the students failed to read any of the suggested reading material, and chose to answer the questions using sentences taken from the lecture notes, without any analysis or synthesis over the various parts of the material. So, answers were quite superficial and repetitive as if they had been memorized. Many students did not bother to answer what was being asked, and just wrote whatever they wanted. Obviously these were not rewarded with marks.

Question 3

Few students took this question, but they did quite well (average mark of 68%). Part (a) was answered correctly in most cases (bookwork), as was part (b). Part (c) was generally OK, with the majority of students illustrating the right context for each of the serialisations.

Part (d) was referring to an application example, and was generally well answered. There were some issues with the casting operator (it has to convert the string in the query, rather than in the RSS).

Finally, in part €, some answers failed to specifically relate to linked data in this context, e.g. what possibilities there are for linking, integration, machine-supported access etc. Weak answers were mostly generic and not really contextualising what the benefits of ontologies or SKOS could be in the given context.