1. General comments

These comments are made in relation to unmoderated, unapproved marks.

21 candidates took the examination. Of these, 10 (47.62%) achieved a First Class mark, 2 (9.52%) an Upper Second, 3 (14.29%) a Lower Second and 4 (19.05%) a Third Class mark. 2 candidates (9.52%) failed the examination, with a mark under 40%. Over half the candidates achieved an Upper Second or First (57.14%). The average mark was 62.14%. The top mark was 86.67%, the bottom mark 25%.

All candidates answered questions 1 and 2; 10 answered question 3; 3 answered question 4; and 6 answered question 5.

2 candidates answered only 2 questions. The bottom mark was for a candidate answering all 3 questions.

Overall, several answers were characterized by the following:
• Avoidance (no answer at all) or major underdevelopment (in relation to marks available) of parts requiring argumentation, discussion, analysis, evaluation and critique.
• An apparent concomitant lack of awareness of major concepts, techniques and trends.
• Apparent ability to describe only, without supplying required examples and/or assessment of issues or impact.
• Vagueness, to the point of leaving the examiner with no option but to decide that the candidate had no understanding of the issue/notion, as nothing concrete was forthcoming.
• Inattention to detail, leading to inconsistencies.
• Lack of justification for some choice or decision, despite this being explicitly required in the question part, for full marks.
• Not answering the question asked.

One may hypothesise that such features of answers point to a lack of engagement with study resources, i.e., little in the way of reading up on topics (even of the relevant chapters in the recommended textbook). In other words, where candidates were unable to move beyond description or fact, it appears that they may have severely limited their chances of obtaining good marks by not assimilating enough from wider study (e.g., via the external resources listed on the course unit web site) to enable them to provide examples, justifications, analyses, discussions, etc. It is demonstration of these “higher” levels of intellectual ability that, after all, typically characterises the graduate who obtains a good degree grade.

One exam script contained extremely poor handwriting throughout, to the point that several words and phrases were illegible. If an examiner cannot read an answer, he cannot award marks.

2. Specific comments

Question 1

The average mark was 14.62/20, with the top mark being 19/20 and the bottom mark 6/20.

Q1a ii: There was little questioning of the methodology of classification of queries.

Q1b was mainly well answered, although several candidates failed to explain their indexing decisions properly, leading to inconsistency in establishment of matrices and inverted indices. Attention to detail is critical. E.g., if one indexes “Earth” and the query is “earth” then there will be no match due to case differences, if no mention of case folding is made. In other words, do not make assumptions. One candidate was unaware of the semantics of Boolean query.

Q1c: Several candidates successfully demonstrated appropriate understanding of the backtracking behaviour when using skip pointers, and had clearly practised applying the algorithm. Others derived the correct skip value but then applied it incorrectly. Some failed to correctly specify the backtracking behavior.

Q1e: Part ii was under-answered by some, given the marks available.

Question 2

The average mark was 12/20, with the top mark being 18/20 and the bottom mark 3/20.

Q2a was well answered by many, but others were unable to do more than describe the concepts without taking account of various decisions in relation to these or the impact of such decisions. Transliteration and translation are not the same thing.

Q2b was relatively poorly answered, that is, the concepts of idf, tf and tf-idf were not well differentiated and explained. There is a very large literature on idf and its cousins.

Q2c was a test of basic understanding that several candidates failed. Some failed to give the required justification, so did not get full marks. Justification in such a question is required to enable the examiner to determine whether or not the candidate is simply taking a yes/no guess.

Q2d was mainly well answered, although several candidates only explained one way (or gave two ways, but their second was the same as the first).
Q2e led to some poor responses demonstrating a lack of knowledge of Euclidean distance.

Q2f was answered well by several, but poorly or not at all by others. It was designed to test understanding of the nature of the Vector Space Model.

Q2g was answered well, although some answers were quite underdeveloped for the marks available.

Question 3

The average mark was 10.2/20, with the top mark being 15/20 and the bottom mark 2/20.

Q3a: several candidates answered only 1 or 2 parts. This question was testing understanding of concepts by couching descriptions in different ways using variant terms.

Q3b was generally well answered, although some candidates simply reproduced the formula for F-measure without actually answering the question, thus the examiner could not tell whether they actually understood the implications of the formula.

Q3c was answered very well by several candidates, who had clearly been practising how to calculate precision and recall in a ranked retrieval system, and how to plot 11-point interpolated precision. Others managed to correctly calculate precision and recall, but then did not correctly calculate the interpolated precision values.

Q3d was a good distinguisher, with few candidates answering correctly.

Q3e was answered well by some, bringing in relevant knowledge and examples. Others under-answered for the marks available, or gave vague, general answers.

Question 4

The average mark was 9.33/20, however only 3 candidates tackled this question, with 2 of the 3 failing and the third obtaining a First for this question.

Q4a: the RDF/XML descriptions were correct, despite a claim to the contrary that revealed a lack of knowledge regarding how RDF triples are expressed in RDF/XML.

Q4b and c: was answered well by only 1 candidate, who clearly explained their reasoning.

Given the small sample of candidates, it is not possible to make firm conclusions about performance on this question, however the overall attendance(?)/assimilation/revision/examination strategy of two of the candidates would appear to have limited their choices in terms of topics they could answer well on.

Question 5

The average mark was 15.67/20, with the top mark being 18/20 and the bottom mark 12/20.

This was a single-part essay question and was in general well tackled by 6 candidates, with interesting, informative, well-argued and supported views and discussions being put forward, demonstrating knowledge of recent developments in the field.

The numbers answering this question and the level of performance were superior to the previous year, which had seen some vague, general, experiential answers.

The topic of this question was formulated to enable candidates to demonstrate their knowledge of the field while doing so via appropriate structured, justified argument and discussion. Moreover, it was linked to coursework themes. Results demonstrated that they were successfully able to integrate knowledge of the various course topics and to give strong evidence of deep learning.

30 September 2013