

---

# 2910325 Data compression

## Examiner's report 2010 – Zone B

---

### Important notice

From 2010 onwards the CIS325 examination paper will consist of just THREE questions. Candidates should attempt all three questions on the paper.

---

### General remarks

CIS325 examination papers cover typically the most important topics of the course material. The questions may be different from year to year, but they can be classified as three types. The first type of questions are bookwork, asking you to explain a term or concept. The second type of questions are similar to what you have seen in the exercises of the textbook, subject guide or what you have done in the coursework, asking you to engage in certain activities, for example, describe a technique, method, known algorithm, to analyse an output, or to explore certain characteristics. The third type questions are unseen, allowing outstanding performance to shine.

As an examination candidate, you are expected to have fully understood issues and mastered the main techniques learned from the subject guide. Excellent candidates are expected to have also developed certain analytical abilities in problem solving. While revising for the examination, you should focus on the bookwork, exercises, coursework and questions in the past examination papers, in that order. You should not try to memorise what you have seen but focus on understanding what you have done and why a certain way does not work.

Your grade depends crucially on how successfully you can demonstrate the required knowledge and skills in the examination. What you write on the exam book is vitally important. It is hence essential for you to be able to present the solutions in a best possible way. For example, the answers on your scripts should be written in logical and coherent steps. You should show the entire process of deriving a solution, not a final result alone. The corresponding question numbers (e.g. Question 1, Question 3) and part numbers (e.g. (a), (b), ii, iii.) should be clearly marked in your answers. The solution to each question should begin on a new page if possible. Your handwriting should be easy to read. You may use pens of any natural dark colour, preferably black or blue, but never red or green which are reserved for marking your exam scripts. It is a good practice sometimes to leave a note for the marker to clarify or highlight things, especially when solutions to some parts of a question are presented in different places. Most importantly, candidates need to read questions **carefully**.

Good timing in an examination is critical for your success. You should therefore quickly draft a schedule at the beginning of the examination to allocate the time for each question. This is to avoid the situation where you spend too much time on one question and do not have enough time for another. A similar timing plan should be made for each sub-question or part of the question.

There are a total of three questions in this year's exam paper. Candidates are required to attempt all the three questions. I am very pleased to see that this seems to have worked very well in terms of the good overall examination results.

We now discuss the questions one by one:

---

## Specific remarks

### Question 1

- a. This part of the question tests candidates' general knowledge about Data compression. Most students failed to give a full answer. It seems that students were too focused on the technical side of the subject and overlooked the other, potentially more interesting, knowledge side of the subject.
- b. A good answer to this part of the question would consist of three sections. The first section would set up the argument and the second section would provide supportive reasoning, with an example. The final section would include the calculation for the percentage. A perfect answer would also conclude or summarise the results.
- c. An easy approach to answer this part of the question is to follow the question setting and answer the two sub-questions one by one. For each sub-question, a good answer would give first a true-or-false answer, and then follow it up with the reasons for the answer. The optimal arguments are closely related to the entropy of the source. The theoretical bound is often used to compare with the average length of the derived code.
- d. A good answer to this part of the question would be threefold: first, a proposal to extend the alphabet. Secondly, derive a canonical Huffman code on the extended alphabet. Finally, compare and show the improvement on the gaps between the equivalent entropy and average length.

### Question 2

- a. A good answer to this part of the question would follow the requirements and answer the two sub-questions accordingly.
  - i. The answer to this section can be obtained by checking the McMillan's inequality.
  - ii. This is an unseen question but the answer can be derived from the knowledge of the Kraft-McMillan's inequality. Note that the length of a variable length code here is referred to, by definition, as the shorter average length of the codewords.
- b. This is, again, an unseen question. A good way to answer this part of the question is to adopt the sub-question numbers and answer each of the sub-questions individually.
  - i. Compute the frequencies of the characters before converting it to the probabilities.
  - ii. Compute the minimum length of the file using the entropy.
  - iii. Note the minimum length of the image segment would remain the same as the probability distribution of the characters if the file remains unchanged.
- c. A good answer to this part of the question would consist of two sections. The first section would be used to explain the predictive rule,

i.e. the relations between the four variables: T, S, Q and x. The second section should give an example to demonstrate your understanding of the rule.

- d. There are three sub-questions in this part of the question. Your answer should be divided into three sections accordingly. In the first section, you should give a yes-or-no answer. This should be followed by a concise explanation in the second section. Finally, a suitable example should be given in the third section.

### Question 3

- a. A good answer to this part of the question would demonstrate how a binary code can be transferred to a Reflected Grey Code. It is essential that you show all your work steps in order to gain the maximum value of credits.
- b. This part of the question requires a flowchart for the adaptive Huffman encoding algorithm. Of course, you need to demonstrate your understanding of the algorithm, not to just give a flowchart. You may therefore add brief explanations if necessary.
- c. A good answer to this part of the algorithm would include two sections. The first section would show the flowchart for LZ77 decoding algorithm. The second section would decode the given tokens under the assumptions on the history buffer and the lookahead buffer. Again, you need to show all your work to avoid losing credits.
- d. This question is straightforward if you know how the arithmetic decoding algorithm works. A good answer would contain two sections as required. The first one would contain a concise explanation of the given four variables as required, and the second section would contain a filled form to demonstrate the decoding steps. Note you should always write the solutions on the scripts unless it is stated otherwise.

---

## Summary

Good performance in the examination depends upon not only your sound knowledge covering all the required topics but also your abilities of demonstrating your knowledge and analytical skills. The importance of understanding the questions in the examination cannot be emphasised more. Candidates are, again, advised to read the questions **carefully**. You should make sure that for each question you fully understand what is required and what parts or sub-questions are involved explicitly and implicitly. You are encouraged to take notes, if necessary, while attempting the questions. Above all, you should be completely familiar with the course materials. Good candidates should be prepared to solve problems in some unseen form by applying their knowledge gained from the course.