2910325 Data compression

Examiner's report 2010 – Zone A

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

This part of the question tests 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.

An easy way to answer this part of the question is to divide your answer into three sections. The first section would discuss the main differences between the lossless and lossy algorithms. The second section would explain the general objectives of the lossy compression. Finally, a suitable example should be given in the final section.

A good answer to this part of the question would consist of two sections. The first section would highlight the differences between a static and adaptive compression system. You should include, as required, a diagram to show the model-coder structure in your discussion. The second section would include a suitable example to support your discussion in the previous section. An excellent solution would also include some explanations on the diagram to highlight the points that you want to make.

There are three sub-questions in this part. You should adopt the question numbers and answer them one by one.

i. Two sections should be included in this part: a fixed length binary code and a Huffman code for the given alphabet.

ii. Comments on the perfect binary trees and complete binary trees.

iii. A good way to answer this part of the question is to first compute, for both codes, the gaps between the average length of the fixed length-code and the entropy; and then the gap between the Huffman code and the entropy. Finally, conclude that the code that leads to a smaller gap is the one that is closer to optimal.

Most candidates this year successfully attempted some parts of the question, but very few provided a full answer. This shows certain weaknesses on the part of the students in mastering the relevant materials.

Question 2

A good answer to this part of the question includes two sections. The first one should show that the Kraft-McMillan's inequality is held, and conclude that the reason why John can find a uniquely decodable binary code. The second section should include an appropriate example to demonstrate your understanding.

There are four sub-sections in this part of the question and you should answer them one by one. It is crucial that all the question, sub-questions and part numbers are marked clearly on your scripts to gain all the credits. All the sub-questions are based on the given segments of an image.

i. A good answer to this part would consist of a process of deriving a canonical minimum-variance Huffman code.

ii. A good demonstration would be expected in this section to show how a binary code can be transformed to the Reflected Grey Code (RGC).

iii. You should demonstrate in this section how the coded image segment can be separated into a number of bit planes in order to achieve a better overall compression ratio. Note that you do need to show how a selected compression/decompression technique actually works, preferably still in steps, instead of just mentioning what might be done.

iv. A good answer to this part of the question would describe how a progressive image compression technique such as pyramid coding may be applied to the given segment.

Question 3

A good answer to this part of the question would simply be a flowchart for encoding in order to outline how the adaptive Huffman algorithm for decoding works.

This part of the question requires a step-by-step trace of the execution of the adaptive Huffman algorithm. A good approach is to simply write on each step the input string, the value of each of the variables 'input', 'output', 'alphabet' and a 'tree diagram', as requested. To avoid mistakes, you may also like to draw a flowchart to outline the encoding algorithms for the adaptive Huffman algorithm.

Similarly, a good solution to this part of the question would consist of a series of steps of execution of the LZW decompression algorithm. These steps should demonstrate the detailed input, output, the content of the dictionary and, finally, a summary of the entire output.

A good answer to this part of the question would consist of two sections. The first section explains the given 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.

This is an unseen question, but essentially it is a bookwork type question asking for a concept called the *Peak Signal To Noise Ratio*. It would be simple if you know the concept. Unfortunately, many students failed to answer this question this year. One possible reason could be that candidates are focused too much on those questions that appeared in the past examination papers.

Summary

Good performance in the examination depends not only upon sound knowledge, covering all the required topics, but also your ability to demonstrate your knowledge and analytical skills. The importance of understanding the questions in the examination cannot be emphasised enough. 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.