THE BCS PROFESSIONAL EXAMINATIONS

Professional Graduate Diploma

April 2006

EXAMINERS' REPORT

Knowledge based Systems

General

Question 1

1. KBS are developed from an understanding of how people solve problems in particular application domains.

a) Describe the process of case-based reasoning. **(8 marks)**

b) Explain the process of analogical reasoning. **(5 marks)**

c) Consider the characteristics of problems, and specify three features that warrant the application of knowledge-based technologies. Illustrate the discussion with one example of a task that demonstrates the need for KBS, and one task for which KBS are unnecessary. **(12 marks)**

Answer Pointers

1a) Case-based reasoning has been formalized as a four-step process:

1. **Retrieve**: Given a target problem, retrieve cases from memory that are relevant to solving it. A case consists of a problem, its solution, and, typically, annotations about how the solution was derived. For example, suppose Fred wants to prepare blueberry pancakes. Being a novice cook, the most relevant experience he can recall is one in which he successfully made plain pancakes. The procedure he followed for making the plain pancakes, together with justifications for decisions made along the way, constitutes Fred's retrieved case.

2. **Reuse**: Map the solution from the previous case to the target problem.
   This may involve adapting the solution as needed to fit the new situation. In the pancake example, Fred must adapt his retrieved solution to include the addition of blueberries.

3. **Revise**: Having mapped the previous solution to the target situation, test the new solution in the real world (or a simulation) and, if necessary, revise.
   Suppose Fred adapted his pancake solution by adding blueberries to the batter. After mixing, he discovers that the batter has turned blue -- an undesired effect. This suggests the following revision: delay the addition of blueberries until after the batter has been ladled into the pan.

4. **Retain**: After the solution has been successfully adapted to the target problem, store the resulting experience as a new case in memory. Fred, accordingly, records his newfound procedure for making blueberry pancakes, thereby enriching his set of stored experiences, and better preparing him for future pancake-making demands

1b) Analogical reasoning [5 marks]
Analogy is an inference or an argument from a particular to another particular, as opposed to deduction, induction and abduction, where at least one of the premises or the conclusion is general. The word *analogy* can also refer to the relation between the source and the target themselves, which is often, though not necessarily, a similarity.

1c) Tasks
We may summarise by noting that such tasks are *knowledge intensive*, which means that methods gained through experience and learning are needed in order to produce a solution. Such *knowledge-based* methods should be efficient when compared to the basic trial and error or brute force approaches in which every possibility is attempted until an acceptable solution is found (if it exists). The latter forms of method do not constitute intelligent behaviour.

A suitable task is one that can be tackled in many different ways, none of which can guarantee a satisfactory solution (but still offer the possibility of a solution) since this kind of task would require some of the characteristics introduced earlier (e.g. judgement and reasoning with uncertainty). Architectural design is a good example of this kind of *open-world* or *ill-defined* problem, which needs *approximate* or *heuristic* reasoning.

Examples of knowledge intensive tasks are:
- Diagnosis of malfunctions or diseases
- Configuration of complex objects such as computer systems or designing offices
- Planning sequences of actions such as scheduling timetables

In contrast, a problem that may be solved by following a step by step procedure (e.g. a formula), which is well determined and guaranteed to produce a solution would not qualify as demanding intelligence: e.g. solving simple algebraic problems in maths, or determining standard statistical procedures. These kinds of problem may be solved *algorithmically* by any computing machine that can follow a procedure.

It should be noted that any task can be tackled using intelligent methods, though, unless there is a significant improvement in efficiency from using intelligent methods, the task should not be admitted as one that demands intelligence.

Furthermore, many tasks can be reduced to the problem of searching for a solution in a search space. Accordingly, much of artificial intelligence work is involved with finding effective ways of structuring and representing the search space, and with generating efficient methods for conducting searches. Tasks that push the limits in either of these areas are good candidates for intelligent KB systems.

Examiner's Comments
This question addressed fundamental issues around the nature of knowledge based decision-making. As such all students attempted it and the majority passed. The average mark was not high though a few students gave fairly full answers and achieved good results. Confusion over the difference between CBR and analogical reasoning may have contributed to lower marks for some students.

Question 2

2. Problem solving involves both inference and searching for a solution.

a) Explain the principle of rule inference and describe both forward and backward chaining using illustrative examples. **(10 marks)**

b) Explain the difference between inductive and deductive reasoning. **(5 marks)**

c) Explain both brute-force and heuristic search methods and discuss their relative merits. **(10 marks)**

Answer Pointers

a) Chaining [10 marks]

Forward chaining starts with the available data and uses inference rules to extract more data (from an end user for example) until an optimal goal is reached. An inference engine using forward chaining searches the inference rules until it finds one where the **If** clause is known to be true. When found it can conclude, or infer, the **Then** clause, resulting in the addition of new information to its dataset.

Inference engines will often cycle through this process until an optimal goal is reached.

For example:
I have a pet named Fritz, he's green and he hops, what is he?
1. **If** Fritz hops - **Then** Fritz is green
2. **If** Fritz is green - **Then** Fritz is a frog

Forward-chaining inference is often called data driven — in contrast to backwardchaining inference, which is referred to as goal driven reasoning.

Backward chaining starts with a list of goals (or a hypothesis) and works backwards to see if there are data available that will support any of these goals. An inference engine using backward chaining would search the inference rules until it finds one which has a **Then** clause that matches a desired goal. If the **If** clause of that inference rule is not known to be true, then it is added to the list of goals (in order for your goal to be confirmed you must also provide data that confirms this new rule).

For example, suppose a rulebase contains two rules and that the goal is to conclude that Fritz is a frog, given that he hops:

1. **If** Fritz hops - **Then** Fritz is green
2. **If** Fritz is green - **Then** Fritz is a frog

This rulebase would be searched and rule (2) would be selected, because its conclusion (the **Then** clause) matches the goal (that Fritz is a frog). It is not yet known that Fritz is green, so the **If** statement is added to the goal list (in order for Fritz to be a frog, he must be green). The rulebase is again searched and this time rule (1) is selected, because its **Then** clause matches the new goal just added to the list (that Fritz is green). The **If** clause (Fritz hops) is known to be true and therefore the goal that Fritz is a frog can be concluded (Fritz hops and therefore must be green, Fritz is green and therefore must be a frog).

Because the list of goals determines which rules are selected and used, this method is called goal driven, in contrast to data-driven forward-chaining inference.

b) Reasoning [5 marks]

Reasoning is the act of using reason to derive a conclusion from certain premises, using a given methodology. The two most commonly used explicit methods to reach a conclusion are deductive reasoning and inductive reasoning.

Types of reasoning
In deductive reasoning, given true premises, the conclusion must follow and it cannot be false. This type of reasoning is non-ampliative - it does not increase one's knowledge base - since the conclusion is inherent to the premises. A classical example of deductive reasoning are syllogisms for example:

- all humans are mortal,
- Socrates is a man,
- therefore, Socrates is mortal.

In inductive reasoning, on the other hand, when the premises are true, then the conclusion follows with some degree of probability. This method of reasoning is ampliative, as it gives more information than was contained in the premises themselves. A classical example comes from David Hume:

- The sun rose to the east every morning up till now,
- therefore the sun will rise to the east also tomorrow.

c) Searching [10 marks]
A searching algorithm requires a target for which to search. The list is searched until either the target is located or the algorithm has determined that the target is not in the list. A comparison must be made to determine if the current element retrieved from the collection is the target one; therefore, a measure of similarity is needed. One of the fields, called the *key* field, serves as the measure on which comparison is performed.

The set of all possible solutions to a problem is called the search space. Brute-force search or uninformed search algorithms use the simplest, most intuitive method of searching through the search space, whereas informed search algorithms use heuristics to apply knowledge about the structure of the search space to try to reduce the amount of time spent searching.

Uninformed search
An uninformed search algorithm is one that does not take into account the specific nature of the problem. As such, they can be implemented in general, and then the same implementation can be used in a wide range of problems thanks to abstraction. The drawback is that most search spaces are extremely large, and an uninformed search (especially of a tree) will take a reasonable amount of time only for small examples. As such, to speed up the process, sometimes only an informed search will do.

Informed search
In an informed search, a heuristic that is specific to the problem is used as a guide. A good heuristic will make an informed search dramatically out-perform any uninformed search.
There are few prominent informed list-search algorithms. A possible member of that category is a hash table with a hashing function that is a heuristic based on the problem at hand. Most informed search algorithms explore trees, such as the Best-first search, which is a search with a heuristic that attempts to predict how close the end of a path is to a solution, so that paths which are judged to be closer to a solution are extended first. Efficient selection of the current best candidate for extension is typically implemented using a priority queue.

Examiner's Comments
This question required understanding of essential techniques for the operation of knowledge based systems. Only half of the candidates attempted the question, perhaps as it may have been seen to be a little technical in nature. That said, the average performance was good with some high marks achieved.

Question 3

3. Development of any software system requires careful management in order to ensure that a quality product is produced. Produce a project management framework that is suitable for the construction of an interactive rule based advisory system by addressing the following tasks:

a) Identify all the main stakeholders involved in the project, and explain their respective roles and responsibilities in the construction of an intelligent knowledge based system. **(8 marks)**

b) Identify a suitable project management and software development methodology, and describe the main phases in the development process as they relate to knowledge engineering in particular. **(10 marks)**

c) Describe the measures that should be taken to ensure quality throughout the entire development process.

(7 marks)

Answer Pointers

3a) Individuals involved with expert systems [8 marks]
There are generally three individuals having an interaction with expert systems. Primary among these is the end-user; the individual who uses the system for its problem solving assistance. In the building and maintenance of the system there are two other roles: the problem domain expert who builds the knowledge base, and a knowledge engineer who assists the experts in determining the representation of their knowledge and who defines the inference technique required to obtain useful problem solving activity. Additionally, other stakeholders are the project manager, who takes
operational control of the project, and the board of directors and clients, who are responsible for commissioning the project.

3b) Project management [10 marks]
Methodologies such as Prince2 or DSDM (agile) could be cited. Their main tasks should be listed, e.g. for Prince2: project startup, project initiation, construction and closure. Each of the main tasks should be discussed from a management perspective, e.g. what should the project manager do during the project initiation phase (assessing risk, planning and scheduling, etc.).

3c) Quality [7 marks]
There are four main elements that make up quality management:
- Quality Management System
- Quality Assurance Function
- Quality Planning
- Quality Control

At least two of these should be described briefly.

Examiner's Comments
This question addressed non-technological management issues relating to the development of KBS. Three quarters of the candidates attempted the question with reasonable results. Some students confused systems development methods with project management methods and consequently focused on the wrong aspects in their answers. However, there were also some good precise answers, particularly to part a.

Question 4

4. Suppose a genetic algorithm uses chromosomes of the form x = abcdefgh with a fixed length of eight genes. Each gene can be any digit between 0 and 9. Let the fitness of individual x be calculated as: $f(x) = (a + b) - (c + d) + (e + f) - (g + h)$, and let the initial population consist of four individuals with the following chromosomes:

x1 = 6 5 4 1 3 5 3 2

x2 = 8 7 1 2 6 6 0 1

x3 = 2 3 9 2 1 2 8 5

x4 = 4 1 8 5 2 0 9 4

a) Evaluate the fitness of each individual, showing all your workings, and arrange them in order with the fittest first and the least fit last. **(8 marks)**

b) Perform the following crossover operations:

i) Cross the fittest two individuals using one–point crossover at the middle point. **(4 marks)**

ii) Cross the second and third fittest individuals using a two–point crossover (points b and f). **(4 marks)**

iii) Cross the first and third fittest individuals (ranked 1st and 3rd) using a uniform crossover. **(4 marks)**

c) Suppose the new population consists of the six offspring individuals received by the crossover operations in the above question. Evaluate the fitness of the new population, showing all your workings. Has the overall fitness improved? **(5 marks)**

Answer Pointers

a) Evaluate the fitness of each individual, showing all your workings, and arrange them in order with the fittest first and the least fit last.

Answer:

$f(x1) = (6 + 5) - (4 + 1) + (3 + 5) - (3 + 2) = 9$

$f(x2) = (8 + 7) - (1 + 2) + (6 + 6) - (0 + 1) = 23$

$f(x3) = (2 + 3) - (9 + 2) + (1 + 2) - (8 + 5) = -16$

$f(x4) = (4 + 1) - (8 + 5) + (2 + 0) - (9 + 4) = -19$

The order is x2, x1, x3 and x4. **[8 Marks]**

Perform the following crossover operations:

i) Cross the fittest two individuals using one–point crossover at the middle point.

Answer: One–point crossover on x2 and x1:

x2 = 8 7 1 2 6 6 0 1

x1 = 6 5 4 1 3 5 3 2

O1 = 8 7 1 2 3 5 3 2

O2 = 6 5 4 1 6 6 0 1 **[4 Marks]**

ii) Cross the second and third fittest individuals using a two–point crossover (points b and f).
Answer: Two–point crossover on x1 and x3

x1 = 6 5 4 1 3 5 3 2

x3 = 2 3 9 2 1 2 8 5)

O3 = 6 5 9 2 1 2 3 2

O4 = 2 3 4 1 3 5 8 5 **[4 Marks]**

iii) Cross the first and third fittest individuals (ranked 1st and 3rd) using a uniform crossover.

Answer: In the simplest case uniform crossover means just a random exchange of genes between two parents. For example, we may swap genes at positions a, d and f of parents x2 and x3:

x2 = 8 7 1 2 6 6 0 1

x3 = 2 3 9 2 1 2 8 5

O5 = 2 7 1 2 6 2 0 1

O6 = 8 3 9 2 1 6 8 5 **[4 Marks]**

b) Suppose the new population consists of the six offspring individuals received by the crossover operations in the above question. Evaluate the fitness of the new population, showing all your workings. Has the overall fitness improved?

Answer: The new population is:

O1 = 8 7 1 2 3 5 3 2

O2 = 6 5 4 1 6 6 0 1

O3 = 6 5 9 2 1 2 3 2

O4 = 2 3 4 1 3 5 8 5

O5 = 2 7 1 2 6 2 0 1

O6 = 8 3 9 2 1 6 8 5

Now apply the fitness function $f(x) = (a+b)-(c+d)+(e+f)-(g+h)$:

$f(O1) = (8 + 7) - (1 + 2) + (3 + 5) - (3 + 2) = 15$

$f(O2) = (6 + 5) - (4 + 1) + (6 + 6) - (0 + 1) = 17$

$f(O3) = (6 + 5) - (9 + 2) + (1 + 2) - (3 + 2) = -2$

$f(O4) = (2 + 3) - (4 + 1) + (3 + 5) - (8 + 5) = -5$

$f(O5) = (2 + 7) - (1 + 2) + (6 + 2) - (0 + 1) = 13$

$f(O6) = (8 + 3) - (9 + 2) + (1 + 6) - (8 + 5) = -6$

The overall fitness has improved. **[5 Marks]**

Examiners' Comments

This question assesses the student's understanding on how genetic algorithm is working. As the question is about genetic algorithm fundamental concepts such Fitness evaluation of a population and crossover operations, most of the students have answered correctly this question. This shows that students have practiced with such concepts.

Question 5

5. Suppose that you have a small kiosk database of purchased items. The available food items in the kiosk are: Coca-cola, Pepsi, Sprite, Budweiser beer, Guinness beer, Estrella chips, Pringles chips and Taffel chips

The database contains the following purchase transactions:

TID	Items bought
1	Coca-cola, Budweiser beer, Pringles chips
2	Coca-cola, Taffel chips
3	Budweiser beer, Pringles chips
4	Pepsi, Budweiser beer, Guinness beer, Estrella chips
5	Sprite, Estrella chips
6	Pepsi, Budweiser beer, Estrella chips
7	Sprite
8	Budweiser beer, Guinness beer, Estrella chips
9	Pepsi, Estrella chips
10	Coca-cola, Pringles chips

a) What kind of rules do you get with confidence threshold 0.0 and support threshold 0.2? **(8 marks)**

b) Create a hierarchy for the food items in the kiosk **(5 marks)**

c) Figure out if you could get more meaningful information using multi-level association rules. **(12 marks)**

Question 5 Marking Scheme

a) We get association rules {BODY} => {HEAD} (confidence, support) with support threshold 0.2 = 20 % and confidence threshold 0.0 = 0 %

{ Estrella } => { Budweiser } (0.60, 0.30)
{ Budweiser } => { Estrella } (0.60, 0.30)
{ Guinness } => { Budweiser } (1.00, 0.20)
{ Budweiser } => { Guinness } (0.40, 0.20)
{ Pepsi } => { Budweiser } (0.67, 0.20)
{ Budweiser } => { Pepsi } (0.40, 0.20)
{ Pringles } => { Budweiser } (0.67, 0.20)
{ Budweiser } => { Pringles } (0.40, 0.20)
{ Pringles } => { Coca-cola } (0.67, 0.20)
{ Coca-cola } => { Pringles } (0.67, 0.20)
{ Guinness } => { Estrella } (1.00, 0.20)
{ Estrella } => { Guinness } (0.40, 0.20)
{ Pepsi } => { Estrella } (1.00, 0.30)
{ Estrella } => { Pepsi } (0.60, 0.30)

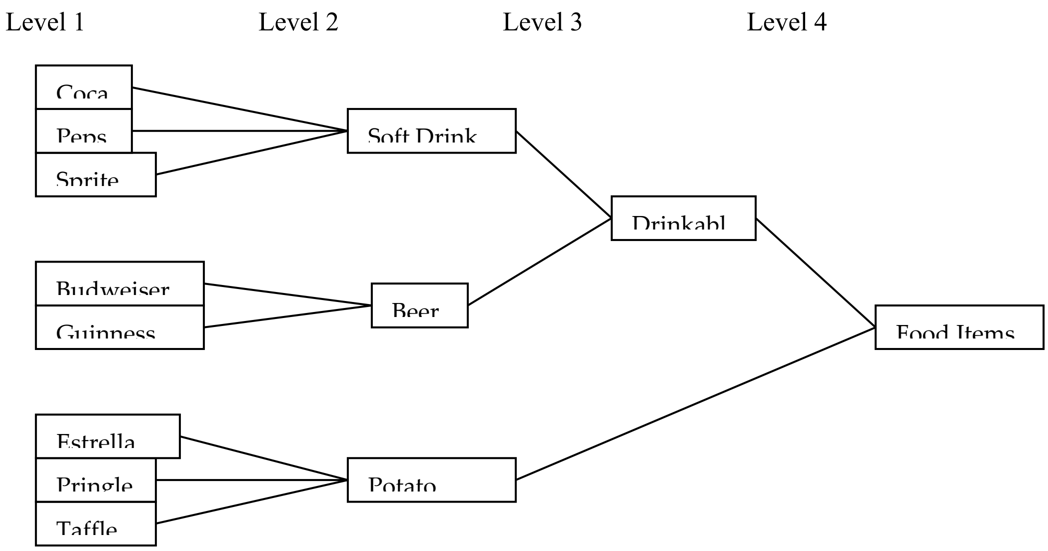
{ Estrella Guinness } => { Budweiser } (1.00, 0.20)
{ Budweiser Guinness } => { Estrella } (1.00, 0.20)

{ Budweiser Estrella } => { Guinness } (0.67, 0.20)
{ Estrella Pepsi } => { Budweiser } (0.67, 0.20)
{ Budweiser Pepsi } => { Estrella } (1.00, 0.20)
{ Budweiser Estrella } => { Pepsi } (0.67, 0.20)

{ Budweiser } => { Guinness Estrella } (0.40, 0.20)
{ Estrella } => { Budweiser Guinness } (0.40, 0.20)
{ Guinness } => { Budweiser Estrella } (1.00, 0.20)
{ Budweiser } => { Estrella Pepsi } (0.40, 0.20)
{ Estrella } => { Budweiser Pepsi } (0.40, 0.20)
{ Pepsi } => { Budweiser Estrella } (0.67, 0.20)

[8 Marks]

b) The hierarchy for the food items in the kiosk is as follows:



[5 Marks]

c) We can figure out more meaningful information using multi-level association rules as presented below using level2, level3 etc….

Purchases for level 2 in the hierarchy
1 soft_drink, beer, potato_chips
2 soft_drink, potato_chips
3 beer, potato_chips
4 soft_drink, beer, potato_chips
5 soft_drink, potato_chips
6 soft_drink, beer, potato_chips
7 soft_drink
8 beer, potato_chips
9 soft_drink, potato_chips
10 soft_drink, potato_chips

We can get more meaningful information using multi-level association rules {BODY} => {HEAD} (confidence, support) for level 2 (purchase)

{ beer } => { potato_chips } (1.00, 0.50)
{ potato_chips } => { beer } (0.56, 0.50)
{ beer } => { soft_drink } (0.60, 0.30)
{ soft_drink } => { beer } (0.38, 0.30)
{ potato_chips } => { soft_drink } (0.78, 0.70)
{ soft_drink } => { potato_chips } (0.88, 0.70)

{ beer potato_chips } => { soft_drink } (0.60, 0.30)
{ beer soft_drink } => { potato_chips } (1.00, 0.30)

{ potato_chips soft_drink } => { beer } (0.43, 0.30)

{ beer } => { potato_chips soft_drink } (0.60, 0.30)
{ potato_chips } => { beer soft_drink } (0.33, 0.30)
{ soft_drink } => { beer potato_chips } (0.38, 0.30)

Purchases for level 3 in the hierarchy
----
1 drinkable, potato_chips
2 drinkable, potato_chips
3 drinkable, potato_chips
4 drinkable, potato_chips
5 drinkable, potato_chips
6 drinkable, potato_chips
7 drinkable,
8 drinkable, potato_chips
9 drinkable, potato_chips
10 drinkable, potato_chips

Again, we can get more meaningful information using multi-level association rules {BODY} => {HEAD} (confidence, support) for level 3 (purchase)

{drinkable} => {potato_chips} (0.90, 0.90)
{potato_chips} => { drinkable } (1.00, 0.90) **[12 Marks]**

Examiner's Comments

This question assesses the student's understanding of data mining concepts. Not all students have answered the last sub question. However, less than a half of students have responded correctly to sub questions (a) and (b).This question is useful for those students with less expressive power in their narrative to gain marks by answering practical questions. There is a need for student to be given more opportunity to practice problem solving using data mining algorithms.