T-79.5102 **Autumn 2006**
Special Course in Computational Logic
Tutorial 10

1. An ornithologist is very interested in observing birds. Every time he sees some moving in the skyline he grabs his binoculars and starts making observations. However, there are also other animals around that are not interesting but draw his attention unnecessarily. To deal with this problem, the ornithologist has developed an AI system to distinguish birds from other animals. The heart of the system is a decision tree based on attributes that get their values as answers to a number of questions:

> Does the animal fly? How many legs has it? Can it swim? Has it fur? What kind of a neck does it have? Does it look like wearing a tail-coat?

The system is trained using the following data set:

Animal	Flies	Legs	Fur	Swims	Neck	Tail-coat	Bird
Starling	Yes	2	No	No	Short	No	**Yes**
Cormorant	Yes	2	No	Yes	Long	No	**Yes**
Giraffe	No	4	Yes	No	Long	No	**No**
Gnat	Yes	6	No	No	Short	No	**No**
Bat	Yes	2	Yes	No	Short	No	**No**
Ostrich	No	2	No	No	Long	No	**Yes**
Penguin	No	2	No	Yes	Short	Yes	**Yes**
Crocodile	No	4	No	Yes	Short	No	**No**
Fluke	No	0	No	Yes	Short	No	**No**

Construct a decision tree for this data set using the expected amount of information as the basis of selecting best attributes. How does the resulting tree classify a human (with values no, 2, no, yes, short, and no)?

2. The personnel manager of an IT company is interviewing candidates for new employees. The table below collects answers of 10 candidates to four different questions posed by the personnel manager. The last column indicates the personnel manager's decision: **Yes**="hired" or **No**="not hired".

	Language proficiency	Programming skills	Education	Working experience	Decision
1	English	Yes	M.Sc.Tech.	No	**Yes**
2	English	No	M.Sc.Econ.	Yes	**No**
3	German	Yes	Engineer	Yes	**Yes**
4	English	No	M.Sc.Tech.	Yes	**Yes**
5	English	No	Merchant	No	**No**
6	French	Yes	M.Sc.Tech.	No	**Yes**
7	English	No	Merchant	Yes	**Yes**
8	German	No	M.Sc.Tech.	No	**No**
9	English	Yes	M.Sc.Tech.	Yes	**Yes**
10	French	Yes	M.Sc.Econ.	No	**No**

(a) Build a decision tree by applying the decision tree learning algorithm to the first six rows of the table. Is the hiring policy applied by the personnel manager exceptional in some sense?

(b) What is the expected information gain (in bits) when the value of the attribute "Language proficiency" is determined (assuming that the value is unknown). Again, use the first six rows of the table.

(c) Test your decision tree using the last four candidates. What is the percentage of correct answers? What kind of changes would result if the decision tree were formed using all candidates?