Write down your name & student number on all answer pages;
Number your answer pages.

# Exam Multimedia Information Systems Master IK

26 October, 2010

**Question 1 (10 points)**

What are *nominal*, *ordinal*, *interval*, *ratio*, *spatial*, and *geophysical* variables?
(10 lines max)

**Question 2 (10 points)**

What is the *semantic gap*? What is the *sensory gap*?
(10 lines max)

**Question 3 (10 points)**

What is *invariance* in a feature space? Give an explanation and a sketch.
(10 lines + 1 fig max)

**Question 4 (10 points)**

A producer intends to make a short video clip, where two people leave each from their own house, happen see each other at the local shop, talk for a while, each pay their goods, and leave again. While initially they were quite unhappy when leaving their houses, after buying some great stuff and talking to each other, they feel much more relieved when coming back home again (well, that's at least the intention of the producer).

a – 5p) Give the temporal relations between the various events

b – 5p) Describe the dimensions of the affective curve and draw an example for the above scenario. (5 lines max)

**Question 5 (10 points)**

Describe the comparison of two text documents using the cosine similarity distance and a vector quantization (bag-of-word) approach. Give all relevant steps and equations where necessary.
(15 lines max)

**Question 6 (15 points)**

What is *Average Precision*? Give a short explanation (5 lines max), the equation, and an example.

**Question 7 (10 points)**

What are the disadvantages of a *tag recommendation system* using tag co-occurrence alone? (5 lines max)

*exam continues on reverse side*

Write down your name & student number on all answer pages;
Number your answer pages.

**Question 8 (15 points)**
An Indian medicine man wants to predict whether it will rain or not. He decides to use the following observations:

1) Clouds: none, fluffy clouds, cumulus clouds.
2) Wind: heavy, calm
3) Birds: flying low, flying high.
4) Reading tea leaves: wave pattern, circle pattern, star pattern

He observes for several days and gets the following:

| *Clouds* | *Tea* | *Birds* | *Wind* | *Rain* |
|---|---|---|---|---|
| Cumulus | Star | high | calm | yes |
| Cumulus | Star | high | heavy | yes |
| None | Star | high | calm | no |
| Fluffy | Wave | high | calm | no |
| Fluffy | Wave | low | calm | no |
| Fluffy | Wave | low | heavy | yes |
| None | Wave | low | heavy | no |
| Cumulus | Circle | high | calm | yes |
| Cumulus | Wave | low | calm | no |
| Fluffy | Circle | low | calm | no |
| Cumulus | Circle | low | heavy | no |
| None | Circle | high | heavy | no |
| None | Star | low | calm | no |
| Fluffy | Circle | high | heavy | yes |

0,357
0,357 $\frac{\log 0,357}{\log 2}$

Questions:
a – 1p) What is the *a-priori* chance on rain?
b – 1p) What is the Entropy of rain?
c – 3p) Calculate the information gain for "Clouds"
d – 10p) Instead of using information gain, a shortcut is simply counting how many errors one gets when always guessing "yes" or "no" for each branch in a decision tree. In that case, one can assess "information gain" as the reduction in the number of errors after inserting a node in the decision tree. Construct the full decision tree for predicting rain/no rain using this simple strategy.
Hint: if you are uncertain about the above described shortcut, you may also use information gain (rather than error counting) to construct the decision tree – although this might take some time. In that case, recall that log2 a = log a / log 2 on a calculator.