**T-61.5060 EXERCISE 5/2005**

In T3 on 9 November 2005 at 12 o'clock.

Recall the notation for the sets of variables: $ABC = \{A, B, C\}$ etc.

1. Consider data consisting of strings, and the class of substring patterns. I.e., $D = \{w_1, \ldots, w_n\}$, where each $w_i$ is a string in some alphabet $\Sigma$. A pattern $p$ is also a string in the same alphabet, and $p$ occurs in $w_i$ if $w_i = xpy$ for some $x$ and $y$. A string pattern $p$ is frequent if $p$ occurs in sufficiently many strings $w_i$.

   Assume that the positive border of the frequent strings is $\{abcd\}$. What is the negative border?

2. In the lectures it was shown that for frequent sets the negative border is set of minimal transversals of complements of the sets in the positive border. Does anything similar hold for string patterns?

3. A boolean formula in *conjunctive normal form* (CNF) is a conjunction of disjunctions, e.g. $(a \vee b \vee c) \wedge (c \vee d)$. A formula in *disjunctive normal form* (DNF) is a disjunction of conjunctions, e.g. $(a \wedge c) \vee (a \wedge d) \vee (b \wedge c)$. A formula is positive, if it does not contain any negations. For each boolean formula in CNF there is an equivalent formula in DNF. Consider the problem of finding an equivalent DNF formula for a given positive CNF formula. What is the relationship of this problem to finding hypergraph transversals?