**T-61.5060 Algorithmic methods in data mining**

**Exercises October 18, 2007**

1. Consider the modification of episodes where an episode occurs only if all the events of the episodes occur consecutively in the sequence, i.e., their are no additional events in between. For example, if the episode is the parallel episode $AB$, then it occurs in the sequence $\cdots CBAD \cdots$, but not it the sequence $\cdots CBCAD \cdots$. Write an algorithm for finding out in how many windows a parallel episode occurs under this constraint. What is the running time of the algorithm?

2. Let the set of variables be $\{A, B, C, D, E\}$. Consider the collection of frequent sets containing sets $ABCD$, $BCE$, and $ADE$ and their subsets. What is the negative border of this collection?

3. A minimal transversal $X$ of a collection $\mathcal{S}$ of sets is a minimal set that intersects all the elements of $\mathcal{S}$. Prove in some more detail that the negative border of a collection of frequent sets is exactly the set of minimal transversals of the complements of the maximal frequent sets.

4. (Border-like concepts from database theory.) Consider a set $U$ of attributes, and a dataset $D$ whose columns are labeled with the elements of $U \cup \{A\}$. We say that a set $X \subseteq U$ determines the attribute $A$, if any two rows $u$ and $v$ from $D$ that have the same value for the attributes in $X$ also have the same value for $A$. That is, if $u[X] = v[X]$, then $u[A] = v[A]$. This is known as a functional dependency, and denoted $X \to A$.

   Given $D$, consider the collection $\mathcal{S}$ all sets $X$ such that $X \to A$. Show that $\mathcal{S}$ is closed under supersets. Let $\mathcal{T}$ be the collection of sets $Y$ such that $Y \to A$ does not hold. Show that $\mathcal{T}$ is closed under subsets. What is the relation between the minimal elements of $\mathcal{T}$ and the maximal elements of $\mathcal{S}$?