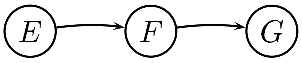
581550 Data mining — tietämyksen muodostaminen
Autumn 2002
Hannu Toivonen

Exercises 3 (due Sep 30–Oct 4)

1. List all subepisodes (with arbitrary partial order, i.e., not only serial subepisodes) of $E \to F \to G$, i.e., of episode



2. Use Algorithm 4.13 to find all parallel episodes in the event sequence of Figure 3.1. Set window width and frequency threshold so that you get useful results.

3. What are the intuitive interpretations of the episode rules
   - if $A \to B$ then $A \to B \to C$
   - if B then $A \to B$
   - if $A \to B$ then $A \to C \to B$
   - if A then B

4. What are the minimal occurrences of the episode $A \to B$ in the following sequence of events? (Time runs on the top, event types are below.)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
A	A	B	A	D	E	B	A	B	B	E	E	A	B	D	A	B	A

   What is the confidence of the rule
   if $A \to B[1.5]$ then $A \to B \to A[3.5]$?

5. Describe informally how the minimal occurrences of $A \to B$ would be found in the above sequence.

6. What kind of minimal episode occurrences can be found in the following sequence fragments
   - ...A...A... vs. ...AA...
   - ...$AAABBB$... vs. ...AB...
   - ...AB...AB... vs. ...A...B......A...B...

   where each position (a letter or a dot) denotes a time point. How well does this solve the problems with windowing (cf. task 6 in exercises 2)?

   Construct examples where the minimal occurrence framework doesn't behave in an intuitive way.