

Adatbányászati algoritmusok

2003/2004. tanév II. félév

VIZSGA FELADATOK – 2004. JÚNIUS 2.

- (1) Írja le Zaki módszerét (**9 pont**). Mitől lehet gyorsabb Zaki módszere az APRIORI módszernél (**3 pont**)?
- (2) Bizonyítsa be, hogy az GSP algoritmus teljes (**12 pont**).
- (3) Ismertesse, és egy példán keresztül mutassa be a MSApriori algoritmus jelöltelőállításának menetét. (**8 pont**).
- (4) Vázzolja mi történik a lenyomat alapú hasonlóságkeresés 3 lépésében (**3 pont**)? Hogyan állítjuk elő a min-hash lenyomatokat (**5 pont**). Mennyire jó ez az eljárás, mit tudunk mondani a helyességéről (**5 pont**)?
- (5) Hogyan definiáljuk egy asszociációs szabály érdekességi mutatóját (**2 pont**)? Milyen hibája van ennek a mutatószámnak (**3 pont**)? Ismertesse azt statisztikai módszert, amivel megállapíthatjuk, hogy egy szabály érdekes-e (**6 pont**).
- (6) Milyen elvárásaink vannak egy jó klaszterező algoritmussal szemben (**5 pont**)? Írja le a single-link működését. Milyen hibái vannak az algoritmusnak (**6 pont**)? Hogyan definiáljuk egy klaszter átmérőjét, illetve a klaszterek közötti távolságot (**3 pont**)?