

# Adatbányászati algoritmusok

2002/2003. tanév II. félév

VIZSGA FELADATOK – 2003. MÁJUS 21.

- (1) Az APRIORI algoritmus melyik hibájára ad megoldást a DHP algoritmus (**2 pont**)? Írja le a DHP működését (**3 pont**)! Mutasson egy konkrét példát arra, amikor a DHP algoritmus jobb az APRIORI-nál (**1 pont**)!
- (2) Előfordulhat-e, hogy az egyszerű mintavételező algoritmus lefutása után, biztosak lehetünk abban, hogy minden gyakori termékhalmazt megtaláltunk? Válaszát indokolja (**3 pont**)!
- (3) Vázzolja a Toivonen algoritmus működését (**5 pont**)! Mutasson példát arra, amikor az algoritmus az adatbázis egyszeri végigolvasásával megtalálja az összes jelöltet és mutasson példát arra is, amikor ehhez két végigolvasás szükséges (**1 pont**). Írja le, hogy kik lesznek a jelöltek ebben az esetben (**1 pont**). Hogyan módosítaná Toivonen algoritmusát, hogy az hierarchikus asszociációs szabályokat is meg tudja találni (**4 pont**)?
- (4) Hogyan definiáljuk a pszeudo-zárt elemhalmazokat (**2 pont**)? Legyen az adatbázisunkban 3 kosár:  $\{A, C\}$ ,  $\{A\}$ ,  $\{A, B, D\}$ . Adja meg a zárt és a pszeudo-zárt termékhalmazokat (**3 pont**)! Bizonyítsa be, hogy tetszőleges termékhalmaznak a támogatottsága megegyezik a lezártjának támogatottságával (**4 pont**).
- (5) Milyen kényszereket vezethetünk be szekvenciális mintákra vonatkozóan (**2 pont**)? Melyik kényszer bevezetése esetén nem lesz igaz az apriori tulajdonság? Az állítását igazolja egy példán keresztül (**2 pont**). Hogyan orvosolja ezt a problémát a GSP algoritmus (**2 pont**)?
- (6) Írja le a hely-érzékeny hash-elésen alapuló hasonló lenyomatpárok keresését (**5 pont**). Mennyire jó ez az eljárás, mit tudunk mondani a hatékonyságáról (**3 pont**)? Jobb-e ez a módszer a futam alapú módszernél (**2 pont**)?
- (7) Hogyan mérhetjük egy közelítő függőség hibáját (**3 pont**)? Hogyan definiáltuk a TANE algoritmusban az attribútumhalmaz szerinti partíciót és a finomított partíciót (**3 pont**)? Hogyan tudjuk eldönteni egy funkcionális függőség érvényességét, ha csak a redukált partíciók adottak (**6 pont**)?
- (8) Milyen elvárásaink vannak egy jó klaszterező algoritmussal szemben (**4 pont**)? Írja le a PAM algoritmus működését. Mitől jobb ez az algoritmus, mint a k-közép (**5 pont**)? Hogyan számítjuk két elem invariáns és a nem invariáns hasonlóságát (**2 pont**)? Bizonyítsa be a kettő közül valamelyik hasonlóságról, hogy metrika (**2 pont**).