

ÓBUDAI EGYETEM

Neumann János Informatikai Kar

Esti Tagozat MSc

FÉLÉVES FELADAT

az Adattárházak építése. c tárgyhoz

NÉV: **Szlaukó András**

NEPTUN KÓD: B7P49U

A DOLGOZAT CÍME:

Webshop adatbázis hasonlósági elemzése

FELADATKIÍRÁS

Egy webshop adatbányászat segítségével szeretné növelni a profitját, ennek érdekében a vásárlók adatbázisában hasonlósági elemzést szeretne elvégezni! A jelenlegi őstábla egy kivonat de jól reprezentálja a fogyasztást és a szokásokat!A pontos feladatokat és az ütemezést a dokumentum feladat ütemezése tartalmazza.

2014. március 27.

.....

Szlaukó András

.....

Nagy István

adjunktus

Eredeti objektumtér „őstábla”

v.sorszám	dvd	laptop	telefon	ipad	laptop táska	alaplapon	merevlemez	hangkártyák	processzorok	memóriák	SSD
vásárló1	+		+			+		+		+	
vásárló2		+			+				+		+
vásárló3		+	+			+	+			+	
vásárló4				+				+			+
vásárló5	+		+		+		+				
vásárló6		+		+		+			+	+	
vásárló7	+	+		+				+			+
vásárló8		+	+	+		+	+		+	+	
vásárló9				+				+			+
vásárló10	+		+		+		+				
vásárló11		+		+		+		+	+	+	+

Feladatok és ütemezés

1.Készítse el a numerikus adatteret az őstáblából!(8.hét)

Z	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
e1	1	0	1	0	0	1	0	1	0	1	0
e2	0	1	0	0	1	0	0	0	1	0	1
e3	0	1	1	0	0	1	1	0	0	1	0
e4	0	0	0	1	0	0	0	1	0	0	1
e5	1	0	1	0	1	0	1	0	0	0	0
e6	0	1	0	1	0	1	0	0	1	1	0
e7	1	1	0	1	0	0	0	1	0	0	1
e8	0	1	1	1	0	1	1	0	1	1	0
e9	0	0	0	1	0	0	0	1	0	0	1
e10	1	0	1	0	1	0	1	0	0	0	0
e11	0	1	0	1	0	1	0	1	1	1	1

2. Alakítsa ki a hasonlósági adatteret / Objektum minősítő paraméterek öntényes beállítása
/Algoritmus módszer használata amivel a redukálendő tér egyre kisebb lesz (9.hét)

Leképezzük a hasonlósági teret az alábbi képlet alapján:

$$\sigma(e_i, e_j) = \frac{\sum_{A_k \in A} \min(Z[e_i, A_k], Z[e_j, A_k])}{\sum_{A_k \in A} \max(Z[e_i, A_k], Z[e_j, A_k])}$$

az e az objektumokat jelöli, A az attribútumokat és Z a numerikus adatteret.

	e1	e2	e3	e4	e5	e6	e7	e8	e9	e10	e11
e1	1	0	0,6	0,3333333	0,3333333	0,6	0,3333333	1	0,4	1	0,75
e2	0	0,4	0,3333333	0,4	0,3333333	0,4	0,6	0,3333333	1	0,6	1
e3	0,5	0,2857143	0,3333333	0,3333333	0,6	0,3333333	0,6	0,75	0,4	1	0,75
e4	0	0,4	0,3333333	0,4	0,1666667	0,75	0,3333333	0,6	0,75	0,6	1
e5	0,3333333	0,1666667	0,6	0,1666667	0,75	0,2	0,75	0,5	0,5	0,75	0,75
e6	0,2857143	0,3333333	0,1428571	0,8	0,3333333	0,3333333	0,6	0,6	0,6	1	0,75
e7	0,1428571	0,6	0,3333333	0,4	0,1666667	0,75	0,3333333	0,6	0,75	0,6	1
e8	0,375	0,5714286	0,25	0,6666667	0,5	0,4285714	0,8	0,6	0,8	1	1
e9	0	0,4	0,3333333	0,4	0,1666667	0,75	0,3333333	0,6	0,75	0,6	1
e10	0,3333333	0,1666667	0,6	0,1666667	0,75	0,2	0,75	0,5	0,5	0,75	0,75
e11	0,25	0,4285714	0,25	0,6666667	0,25	0,6666667	0,4285714	0,6666667	0,8	0,8	1

Objektum minősítő paraméterek önkényes beállítása:

Határhasonlóság: $\sigma_{\min}=0,6$

$\sigma_{\min}=0,6$ feletti hasonlóságok színnel vannak jelölve a táblázatban)

Minimális hasonlósági osztály mérete: $\mu_{\min}=3$

Minimális hasonlósági lefedettség: $C_{\min}=0,6$

	e1	e2	e3	e4	e5	e6	e7	e8	e9	e10	e11
e1	1	0	0,6	0,3333333	0,3333333	0,6	0,3333333	1	0,4	1	0,75
e2	0	0,4	0,3333333	0,4	0,3333333	0,4	0,6	0,3333333	1	0,6	1
e3	0,5	0,2857143	0,3333333	0,3333333	0,6	0,3333333	0,6	0,75	0,4	1	0,75
e4	0	0,4	0,3333333	0,4	0,1666667	0,75	0,3333333	0,6	0,75	0,6	1
e5	0,3333333	0,1666667	0,6	0,1666667	0,75	0,2	0,75	0,5	0,5	0,75	0,75
e6	0,2857143	0,3333333	0,1428571	0,8	0,3333333	0,3333333	0,6	0,6	0,6	1	0,75
e7	0,1428571	0,6	0,3333333	0,4	0,1666667	0,75	0,3333333	0,6	0,75	0,6	1
e8	0,375	0,5714286	0,25	0,6666667	0,5	0,4285714	0,8	0,6	0,8	1	1
e9	0	0,4	0,3333333	0,4	0,1666667	0,75	0,3333333	0,6	0,75	0,6	1
e10	0,3333333	0,1666667	0,6	0,1666667	0,75	0,2	0,75	0,5	0,5	0,75	0,75
e11	0,25	0,4285714	0,25	0,6666667	0,25	0,6666667	0,4285714	0,6666667	0,8	0,8	1

3. 1-es és 2-es lépés végrehajtása(kijelölöm a reprezentatív tér azon objektumát, melynek határhasonlóságú hasonlósági környezetéhez a legnagyobb összesített hasonlóság tartozik az S

térben. Ha több ilyen is van akkor közülük azt kell kiválasztani közülük amelyik a legnagyobb hasonlósági környezettel rendelkezik. (10.hét)

4. 3-as és 4-es pont (ha az előzőekből több is van akkor egy olyat kell kiválasztani közülük amelyik a legnagyobb hasonlósági értékekkel rendelkezik a saját környezetének valamelyik elemével. (11.hét)

5 A hasonlósági osztályozás eredménye /Készítse el a hasonlósági tér táblázatos ábrázolását (12.hét)

6 Hasonlósági tér ábrázolása hasonlósági gráfként / határdevidencia önkéntes beállítása (13.hét)

7. Az elkészült feladat leadása és eredmény értelmezése, mit is jelent ez? Milyen csoportokba sorolhatóak a vásárlók?.A jelenlegi fogyasztást és csoportokat megvizsgálva melyik termék árát érdemes növelni? (14.hét)