

Automatsko indeksiranje i diskriminativna vrednost termina

Zadaci

8.1.1 Odrediti Dajsov indeks sličnosti za sledeće dokumente, predstavljene frekvencijom reči:

1. $D_1 = (1, 3, 2, 0, 5, 1, 0, 0)$ $D_2 = (1, 2, 1, 1, 0, 0, 3, 8)$

2. $D_1 = (1, 2, 0, 3, 1, 0)$ $D_2 = (1, 1, 3, 0, 2, 2)$

8.1.2 Neka je data kolekcija dokumenata sa pridruženim terminima:

$$D_1 \quad T_1, T_2$$

$$D_2 \quad T_2, T_3, T_4$$

$$D_3 \quad T_1, T_3, T_5$$

$$D_4 \quad T_3, T_4, T_5$$

1. Izračunati gustinu prostora Q_{pre_6} ove kolekcije dokumenata koristeći Dajsov indeks sličnosti između dokumenata.
2. Neka se termin T_6 dodeli dokumentima D_2 i D_4 . Odrediti gustinu prostora Q_{posle_6} posle ove dodele i diskriminativnu vrednost $dv_6 = Q_{pre_6} - Q_{posle_6}$ termina T_6 .

8.1.3 Neka je data kolekcija dokumenata sa pridruženim terminima:

$$D_1 \quad T_2, T_3$$

$$D_2 \quad T_1, T_3, T_5$$

$$D_3 \quad T_1, T_2, T_3$$

$$D_4 \quad T_2, T_4, T_5$$

1. Izračunati gustinu prostora Q_{pre} ove kolekcije dokumenata koristeći Dajsov indeks sličnosti između dokumenata.
2. Neka se termin T_6 dodeli dokumentima D_1 i D_3 . Odrediti gustinu prostora Q_{posle_6} posle ove dodele i diskriminativnu vrednost $dv_6 = Q_{pre} - Q_{posle_6}$ termina T_6 .