

Zad 1. Dla zbioru 16 dokumentów opisanych następująco:

X1 = ABEF

X2 = ABEFG

X3 = ABFG

X4 = ACE

X5 = AEH

X6 = BCF

X7 = ABE

X8 = ABF

X9 = AB

X10 = EF

X11 = ABCE

X12 = ABFGH

X13 = ABCEF

X14 = ABCD

X15 = EFGH

X16 = DGH

dokonaj podziału na grupy algorytmem Rocchia. Przyjmij następujące parametry: $N_1 = 2$, $N_2 = 6$, $P_1 = 0,65$, $P_2 = 0,55$. W drugiej iteracji (uwzględniającej wstępny centroid) należy przyjąć parametry: $N_1 = 3$, $N_2 = 6$, $P_1 = 0,5$, $P_2 = 0,45$.

Do określenia współczynnika korelacji należy użyć funkcji:

$$f(x_i, x_j) = \frac{\text{card}(p_i \cap p_j)}{\text{card}(p_i \cup p_j)}$$

gdzie p_i i p_j - zbiory pojęć opisujących dokumenty x_i oraz x_j .

Zad2. W jednym z kroków wiązania w grupy w/g algorytmu Rocchia otrzymano następujące współczynniki korelacji obiektów $x_1 \dots x_6$ z przypuszczalnym centrum grupy:

$k(x_1, x_c) = 0,20$

$k(x_2, x_c) = 0,11$

$k(x_3, x_c) = 0,30$

$k(x_4, x_c) = 0,48$

$k(x_5, x_c) = 0,5$

$k(x_6, x_c) = 0,05$

$k(x_7, x_c) = 0,28$

$k(x_8, x_c) = 0,41$

Założmy, że test gęstości zakłada, iż co najmniej $N_1=6$ obiektów ma współczynnik korelacji (z centrum grupy) wyższy niż $p_1=0,1$ oraz, że co najmniej $N_2=3$ ma ten współczynnik wyższy niż $p_2=0,4$. Jaki jest rezultat tego etapu wiązania obiektów w grupy?

Zad 3. Dla zbioru 8 dokumentów opisanych następująco:

X1 = ACD

X2 = ADF

X3 = DEG

X4 = BCF

X5 = BCG

X6 = ABC

X7 = DEF

X8 = EFG

wykonaj jedną iterację algorytmu Doyle'a. Przyjmij następujący początkowy przydział do grup:

$S_1 = \{X_1, X_2, X_3\}$, $S_2 = \{X_4, X_5, X_6\}$, $S_3 = \{X_7, X_8\}$. Współczynnik skalujący $\alpha = 0,7$.

Zad 4. Przedstaw graficznie reprezentację pni i grup w wyszukiwaniu strukturalnym. Masz do dyspozycji:

P1 = [1,3,3,5,2,1]

P2 = [1,2,0,4,1,6]

P3 = [0,0,0,6,5,4]

Poziom pni

Oraz:

G11 = [1,4,3,5,2,1]

G12 = [0,4,4,0,2,0]

G21 = [0,3,4,3,1,2]

G22 = [0,2,2,4,6,5]

G23 = [1,2,1,3,6,0]

G31 = [0,0,1,4,5,0]

G32 = [1,0,0,4,5,2]

Poziom grup

Omów sposób wyszukiwania dla pytania Q=[1,2,1,3,6,6]. Wykorzystaj w tym celu wzór na **korelację nakładania** w omawianym procesie wyszukiwania.