Mateusz Bulanda - Gorol 245089

① • $N = 4$
• $SUM = (-2, 0, -1)$
• $SUMSQ = (2, 2, 3)$

Centroid: $\left(-\frac{2}{4}, \frac{0}{4}, \frac{-1}{4}\right)$

②

$S_1 = \{a, b, c, g\}$
$S_2 = \{a, b, e, g\}$
$S_3 = \{a\ b\ c\ f\}$

	S_1	S_2	S_3
a	1	1	1
b	1	1	1
c	1	0	1
e	0	1	0
f	0	0	1
g	1	1	0

③

$S_1 = \{(M, a, k), (a, k, k), (k, k, a)\}$
$S_2 = \{(P, a, k), (a, k, k), (k, k, a)\}$

$J(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} = \frac{2}{4} = \frac{1}{2}$

④

*	xy	y	yc	yx	xx
xy	0	1	2	1	1
y	1	0	1	1	2
yc	2	1	0	1	2
yx	1	1	1	0	1
xx	1	2	2	1	0

max = diameter

$\Downarrow$

$\boxed{d = 2}$

(5)

f_i ■ : $(5; 10; 0{,}5; 0{,}6)$ - sensitive function

$f = [\, f_1(x_1) = f_1(x_2)$ OR $f_2(x_1) = f_2(x_2) \,]$

f: ■ $(5; 10; 0{,}75; 0{,}84)$ - sensitive function

(6)

$|A_1| = 10$ $\quad x \in A_1 \wedge x \in A_2 \Rightarrow |A_1 \cap A_2| \neq 0$

$|A_2| = 20$

$$J(A_1, A_2) = \frac{|A_1 \cap A_2|}{|A_1 \cup A_2|}$$

$|A_1 \cap A_2|_{min} = 1$ $\quad |A_1 \cap A_2|_{max} = 10$

$|A_1 \cup A_2|_{min} = 30$ $\quad |A_1 \cup A_2|_{max} = 30$

$$J_{min}(A_1, A_2) = \frac{|A_1 \cap A_2|_{min}}{|A_1 \cup A_2|} = \frac{1}{30}$$

$$J_{max}(A_1, A_2) = \frac{|A_1 \cap A_2|_{max}}{|A_1 \cup A_2|} = \frac{10}{30} = \frac{1}{3}$$

(7)

~~A - chosen possition in both texts are the same~~

~~$P(A) = 0{,}9$~~

A - there is the same character on the randomly choosen possition

$P(A) = 0{,}9$

B - there are the same characters on 20 randomly choosen possitions

$P(B) = 0{,}9^{20} = 0{,}1215$

false acceptance rate = 0,1215

(8) $$TF.IDF(i,j) = \frac{f_{i,j}}{\sum_k f_{k,j}} \cdot \log_2\left(\frac{N}{n_i}\right) > 0$$

$\frac{f_{i,j}}{\sum_k f_{k,j}} > 0$ - i-th term appears in j-th text at least once

$\log_2\left(\frac{N}{n_i}\right) > 0 \Rightarrow \frac{N}{n_i} > 1 \Rightarrow N > n_i$ -

- i-th term doesn't appear in every of N texts