

1)Postings List: 42, 59, 81, 85, 99, 300, 303

Gap List: 42, 17, 22, 4, 14, 201, 3

Ασυμπίεστο μέγεθος: 224 bits.

γ-code: 11111001010 111100001 111100110 11000 1110110 111111101001001 101

Συμπίεση=  $1-59/224=73,7\%$ .

δ-code:11010101010 110000001 110000110 10000 101110 110111001001 1

Συμπίεση= $1-53/224=76,3\%$ .

Variable byte code:10101010 10010001 10010110 10000100 10001110  
0000000111001001 10000011

Συμπίεση= $1-64/224=71,43\%$ .

2)α)127 (11111111)

β)2.097.151 (011111110111111111111111)

γ)Θα χρειαστούν 13 bytes για την κωδικοποίηση του gaps list.

3)

111000111010101111110110111110111

111000111010101111110110111110111

Gap list: <9,6,3,59,15>

Postings list: <9,15,18,77,92>

5)

Term	Doc1	Doc2	Doc3
car	0.67	0.39	0.6
auto	0.51	0.73	0
insurance	0	0.57	0.61
best	0.54	0	0.52

Γνωρίζουμε ότι:

$W=(1+\log(tf)) \times idf$  ο λογάριθμος έχει βάση το 10.

Doc1:

car:  $(1 + \log(28)) * 1.65=(1+1.45) * 1.65=4.04$

auto:  $(1+\log(3)) * 2.08 =(1+0.48)*2.08=3.08$

insurance: 0

best:  $(1+\log(14)) * 1.5 = (1+1.15) * 1.5=3.23$

Doc2:

car:  $(1+\log(5)) * 1.65=(1+0.7) * 1.65=2.81$

auto:  $(1+\log(33)) * 2.08 = (1+1.52) * 2.08=5.24$

insurance:  $(1+\log(33)) * 1.62 = (1+1.52) * 1.62=4.08$

best:0

Doc3:

car:  $(1+\log(24))*1.65=(1+1.38) * 1.65=3.93$

auto:0

insurance:  $(1+\log(29)) * 1.62=(1+1.46) * 1.62=3.99$

best:  $(1+\log(18)) * 1.5=(1+1.26) * 1.5=3.39$

**Normalization:** Διαιρούμε την κάθε γραμμή με το δικό της μήκος.

**Μήκη**

Doc1:  $\text{sqrt}(4.04^2 + 3.08^2 + 3.23^2) = 6.02$

Doc2:  $\text{sqrt}(2.81^2 + 5.24^2 + 4.08^2)=7.21$

Doc3:  $\text{sqrt}(3.93^2 + 3.99^2 + 3.39^2)=6.55$

Τα αποτελέσματα φαίνονται στον παραπάνω πίνακα.

6)α)

Term	query	Doc1	Doc2	Doc3	Prod1	Prod2	Prod3
car	1	0.67	0.39	0.6	0.67	0.39	0.6
auto	0	0.51	0.73	0	0	0	0
insurance	1	0	0.57	0.61	0	0.57	0.61
best	0	0.54	0	0.52	0	0	0
score					0.67	0.96	1.21

Για τα Prod απλά μεταφέρουμε τις γραμμές που έχουν 1 στο query δεξιά.

Score1=  $0.67 + 0 + 0 + 0 = 0.67$

Score2=  $0.39 + 0 + 0.57 + 0 = 0.96$

Score3=  $0.6 + 0 + 0.61 + 0 = 1.21$

Doc3>Doc2>Doc1

β) Κανονικοποίηση idf:

normalization(idf):  $\sqrt{1.65^2 + 2.08^2 + 1.62^2 + 1.5^2} = 3.45$

$$1.65/3.45 = 0.48$$

$$2.08/3.45 = 0.6$$

$$1.62/3.45 = 0.47$$

$$1.5/3.45 = 0.43$$

Prod1:

$$\text{car}: 0.48 * 0.67 = 0.32$$

$$\text{auto}: 0.6 * 0.51 = 0.31$$

$$\text{insurance}: 0.47 * 0 = 0$$

$$\text{best}: 0.43 * 0.54 = 0.23$$

$$\text{score1}: 0.32 + 0.31 + 0 + 0.23 = 0.86$$

Prod2:

$$\text{car}: 0.48 * 0.39 = 0.19$$

$$\text{auto}: 0.6 * 0.73 = 0.44$$

$$\text{insurance}: 0.47 * 0.57 = 0.27$$

$$\text{best}: 0.43 * 0 = 0$$

$$\text{score2}: 0.19 + 0.44 + 0.27 + 0 = 0.9$$

Prod3:

$$\text{car}: 0.48 * 0.6 = 0.29$$

$$\text{auto}: 0.6 * 0 = 0$$

$$\text{insurance}: 0.47 * 0.61 = 0.29$$

$$\text{best}: 0.43 * 0.52 = 0.22$$

$$\text{score3}: 0.29 + 0 + 0.29 + 0.22 = 0.8$$

Term	query	Doc1	Doc2	Doc3	Prod1	Prod2	Prod3
car	0.48	0.67	0.39	0.6	0.32	0.19	0.29
auto	0.6	0.51	0.73	0	0.31	0.44	0
insurance	0.47	0	0.57	0.61	0	0.27	0.29
best	0.43	0.54	0	0.52	0.23	0	0.22
score					0.86	0.9	0.8

Άρα Doc2>Doc1>Doc3