

Exo 10.2

① calculate idf

$N = 3$

term	df	idf
two	1	0,47
tea	2	0,17
Me	2	0,17
you	2	0,17

②

Doc 1	tf_{ran}	tf_w	idf	VF	VF_{norm}
two	2	2	0,47	0,94	0,94
Me	2	2	0,17	0,34	0,34

Smart notation:
NTC.NNN

$\frac{0,94}{0,99} = 0,94$.

length = $\sqrt{(0,94)^2 + (0,34)^2} = 0,99$

query	tf_{ran}	tf_w	idf	VF
tea	1	1	1	1
me	1	1	1	1

"tea me"

Similarity(q, Doc_1) = $0,34 * 1 = 0,34$.

Correction ID N^o 04

Exo n^o 1

① Calculat idf

$$idf_{term} = \log_{10} \frac{N}{df_{term}}$$

Where N = 806791

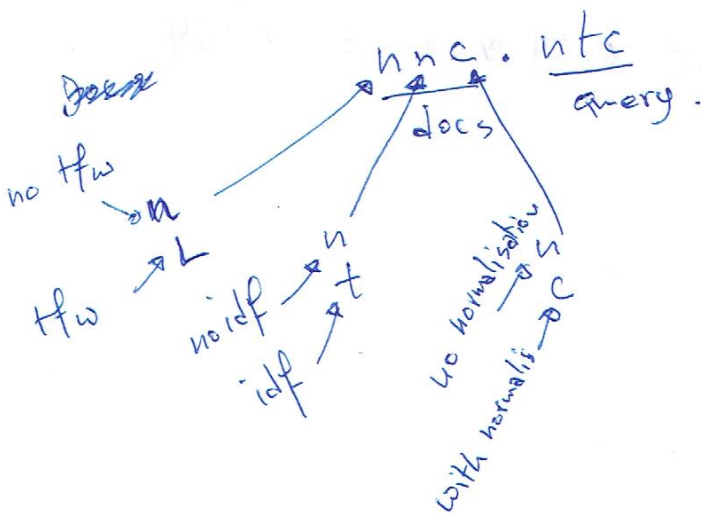
term	df _t	idf _t
car	18165	1,64
Auto	6723	2,107
insurance	19241	1,62
best	25235	1,50

② Calculate similarity of query "car car insurance"

→ Smart notation nnc . ntc

Doc 1	tf _{raw}	tf _w	idf	VF	VF _{norm}
car	12	12	1	12	12/15,74 = 0,76
auto	2	2	1	2	2/15,74 = 0,12
best	10	10	1	10	10/15,74 = 0,63

$$\text{length}_{doc1} = \sqrt{(12)^2 + (2)^2 + (10)^2} = 15,74$$



Query	tf _{raw}	tf _w	idf	VF	VF _{norm}
car	2	2	1,64	3,28	0,189
insur	1	1	1,62	1,62	0,144

$$\text{Length}_{query} = \sqrt{(3,28)^2 + (1,62)^2} = 3,65$$

→ do the same calculation for the rest of docs.

"car" in Doc 1 "car" in query

$$\text{Similarity}(Q, \text{Doc 1}) = (0,76 * 0,189) = 0,14364$$