

TD04 - Information Retrieval and Algorithms

Third Year ISIL - University of M'Sila

Exercise 01

You have the following values of tf of three (3) documents from a collection of 806791 documents, and the values of df of four (4) terms:

Term	Doc1	Doc2	Doc3
car	12	25	39
auto	2	33	1
insurance	0	31	20
best	10	0	12

Table 1: Term Frequencies

Term	df _t
car	18165
auto	6723
insurance	19241
best	25235

Table 2: Document Frequencies

- First, calculate the values of idf?
- Compute the two top scoring documents on the query “car car insurance” for the following SMART notation scheme:
 1. nnc.ntc
 2. ltc.lnc

Exercise 02

Considering the following table of count vector (Tfraw) of a 3 documents and query:

Term	Doc1	Doc2	Doc3
Two	2	0	0
Tea	2	2	0
Me	0	1	2
You	0	1	2

Table 3: Term Frequencies

Term	Query
Two	0
Tea	1
Me	1
You	0

Table 4: Query Vector

- Calculate the idf (inverse document frequency) for each word represented in the three documents.
- What is the result of the query (Tea me) using NTC.NNN smart notation?