

Exam 1 - Computation of TF-IDF

The words car, auto, best should have the following frequencies in the ten documents Doc1, ..., Doc10. Simply assume that all other words in the documents are stop words.

term \ doc	Doc1	Doc2	Doc3	Doc4	Doc5	Doc6	Doc7	Doc8	Doc9	Doc10
car	3	0	0	5	12	0	0	2	8	1
auto	8	6	0	12	0	0	9	1	3	10
best	0	1	7	0	1	5	12	0	2	0

Compute the tf-idf weight for each term using the equations in the slides 14-16.

Exam 2 - Inverted Lists

Assume the following documents

doc-id\text	
Doc1	“italy is world champion 2006”
Doc2	“germany and italy played each other in the semifinal”
Doc3	“germany was in the semifinal 2006”
Doc4	„germany won the semifinal in italy 1990”

Assume that the following terms are stop words: is, and, in, the, was, each, other.

Construct an inverted index according to the figure in slide 25.