

CS 201 Data Structures II – Spring 2024

Instructor: Dr. Muhammad Mobeen Movania

Quiz 7 - Solution

Name: _____

Date: _____

Regn. No. : _____

Time limit: 30 mins

Marks assigned to the question are mentioned at the start of each question.

Q1) (10 marks) Consider the following idf values.

term	df_t	idf_t
car	18,165	1.65
auto	6723	2.08
insurance	19,241	1.62
best	25,235	1.5

The term frequencies for three documents, Doc1, Doc2 and Doc3 are given in the following table:

	Doc1	Doc2	Doc3
car	27	4	24
auto	3	33	0
insurance	0	33	29
best	14	0	17

(a) (5 marks) Generate the inverted index for the terms?

Term	Document Frequency	Postings
auto	6723	1→2→...
best	25235	1→3→...
car	18165	1→2→3→...
insurance	19241	2→3→...

(b) (5 marks) Calculate the tf-idf weights for the terms car, auto, insurance, best, for each document?

$tf_idf_{car, doc1} = 27 * 1.65 = 44.55$ $tf_idf_{auto, doc1} = 3 * 2.08 = 6.24$ $tf_idf_{insurance, doc1} = 0$ $tf_idf_{best, doc1} = 14 * 1.5 = 21$	$tf_idf_{car, doc2} = 4 * 1.65 = 6.6$ $tf_idf_{auto, doc2} = 33 * 2.08 = 68.64$ $tf_idf_{insurance, doc2} = 33 * 1.62 = 53.46$ $tf_idf_{best, doc2} = 0$	$tf_idf_{car, doc3} = 24 * 1.65 = 39.6$ $tf_idf_{auto, doc3} = 0$ $tf_idf_{insurance, doc3} = 29 * 1.62 = 46.98$ $tf_idf_{best, doc3} = 17 * 1.5 = 25.5$
--	---	---