

# Introduction to Data Science.

## Assignment 2: Simple In Database Text

Description	Query	Answer
<b>Problem 1: Inspecting the Reuters Dataset; Basic Relational Algebra</b>		
(a) select:	select count(*) from frequency where docid='10398_txt_earn';	138
(b) select project:	select count(*) from frequency where docid='10398_txt_earn' and count=1;	110
(c) union:	<b>This is the Correct answer:</b> select count(*) from (select * from frequency where docid='10398_txt_earn' and count=1 union select * from frequency where docid='925_txt_trade' and count=1);	335
(c) union:	<b>This is the Grader's Correct answer:</b> select count(distinct term) from (select * from frequency where docid='10398_txt_earn' and count=1 union select * from frequency where docid='925_txt_trade' and count=1);	324
(d) count:	select count(docid) from frequency where term='parliament';	15
(e) big documents:	select count(*) from (select docid from frequency group by docid having sum(count)>300);	107
(f) two words:	select count(*) from ((select docid from frequency where term='transactions' ) as a join (select docid from frequency where term='world') as b on a.docid=b.docid);	3

**Problem 2: Matrix Multiplication in SQL**

<b>(g) multiply:</b>	select value from (select A.row_num,B.col_num, sum(A.value*B.value) as value from A,B where A.col_num=B.row_num group by A.row_num,B.col_num) where row_num=2 and col_num=3;	2874
----------------------	--	------

**Problem 3: Working with a Term-Document Matrix**

<b>(h) similarity matrix:</b>	select max(count) from (select sum(a.count*b.count) as count from (select docid, term, sum(count) as count from frequency where docid='10080_txt_crude' group by docid,term) as a join (select docid, term, sum(count) as count from frequency where docid='17035_txt_earn' group by docid,term) as b on a.term=b.term);	19
-------------------------------	--	----

<b>(i) keyword search:</b>	select max(count) from (select docid, sum(count) as count from (select docid, sum(count) as count from frequency where term='washington' group by docid union select docid,sum(count) as count from frequency where docid='treasury' group by docid union select docid, sum(count) as count from frequency where term='taxes' group by docid) group by docid);	6
----------------------------	--	---