**Problem 1: Inspecting the Reuters Dataset; Relational Algebra**

**(a) select:** Write a query that is equivalent to the following relational algebra expression.

σdocid=10398\_txt\_earn(frequency)

What to turn in: Run your query against your local database and determine the number of records returned as described above. In your browser, turn in a text file, select.txt, which states the number of records.

**(b) select project:** Write a SQL statement that is equivalent to the following relational algebra expression.

πterm( σdocid=10398\_txt\_earn and count=1(frequency))

What to turn in: Run your query against your local database and determine the number of records returned as described above. In your browser, turn in a text file, select\_project.txt, which states the number of records.

**(c) union:** Write a SQL statement that is equivalent to the following relational algebra expression. (Hint: you can use the UNION keyword in SQL)

πterm( σdocid=10398\_txt\_earn and count=1(frequency)) U πterm( σdocid=925\_txt\_trade and count=1(frequency))

What to turn in: Run your query against your local database and determine the number of records returned as described above. In your browser, turn in a text file, union.txt, which states the number of records.

**(d) count:** Write a SQL statement to count the number of documents containing the word “parliament”

What to turn in: Run your query against your local database and determine the count returned as described above. In your browser, turn in a text file, count.txt, which states the count.

**(e) big documents** Write a SQL statement to find all documents that have more than 300 total terms, including duplicate terms. (Hint: You can use the HAVING clause, or you can use a nested query. Another hint: Remember that the count column contains the term frequencies, and you want to consider duplicates.) (docid,term\_count)

What to turn in: Run your query against your local database and determine the number of records returned as described above. In your browser, turn in a text file, big\_documents.txt, which states the number of records.

**(f) two words:** Write a SQL statement to count the number of unique documents that contain both the word 'transactions' and the word 'world'.

What to turn in: Run your query against your local database and determine the number of records returned as described above. In your browser, turn in a text file, two\_words.txt, which states the number of records.