Inf1B Data and Analysis
Tutorial 7 (week 9)

4 March 2010

- Please answer all questions on this worksheet in advance of the tutorial, and bring with you all work. Tutorials cannot function properly unless you do the work in advance.
- Data & Analysis tutorial exercises are not assessed, but are a compulsory and important part of the course. If you do not do the exercises then you are unlikely to pass the exam.
- Attendance at tutorials is obligatory; please let your tutor know if you cannot attend.
- *Background Reading:* Lecture slides on Information Retrieval.

Introduction

In this tutorial we will work on Information Retrieval. Note that the workload for this tutorial is lighter than usual.

1 Information Retrieval

You are looking for a document on `Economic Recession in Scotland` in a huge corpus of documents. Incidentally, you decide to search using the terms: **economy**, **recession**, **Scotland**, **banks** and **business** using an *information retrieval system* and you find three possible documents. You are given the frequency of each of the terms in each document, as shown below:

Terms	economy	Scotland	recession	banks	business
Document 1	10	8	0	2	1
Document 2	0	0	9	9	8
Document 3	2	2	4	4	6
Query	1	1	1	1	1

(a) Which measure would you use with this information to determine which of the 3 documents is the best match for the query?

(b) Compute the measure provided in (a) for all three documents.

(c) Based on your results of (b), which document is the best match? Why?

(d) Do you agree with the results of this analysis? What are the strengths and weaknesses of the measure you used?