## Representing and Accessing [Textual] Digital Information (COMS/INFO 630), Spring 2006 2/28/06: Lecture guides for 2/9/06 (lecture five)

Authors: Matt Connolly and Danish Mujeeb (first 8 pages); Asif-ul Haque and Benyah Shaparenko (subsequent pages).

Some additional questions on the topic of the lecture are below. (Not to say that there's anything wrong with the questions posed in what's attached; these are just other thoughts.)

- 1. Show that there are "somewhat" reasonable assumptions one can impose that allow us to interpret the ratio-of-Poisson-parameters quantity in our isolated-Poisson model as a type of IDF.
- 2. The questions in both sets of lecture guides point out a problem with the  $\log ((N N^{(j)})/N^{(j)})$  formulation of the IDF that results from the Croft and Harper estimate for the binary-attribute RSJ model. Can appropriate "fixes" be (more or less) rigorously derived, perhaps via different assumptions?