

SIMS 202 – Extra Credit
Due December 9, 2004
40 points total (each question is 10 points)

Your response to each question should be a short essay of about one page answering the question. Be sure to include responses to each part of multipart questions.

1. Researchers (and users) over the years have found many problems with Boolean information retrieval systems. Describe and discuss some of the problems with Boolean IR systems and how are these problems addressed in systems using ranked retrieval (such as in vector IR systems)?
2. We know that document collection statistics tend to exhibit Zipf-like distributions. Describe the Zipf distribution and describe and discuss the major significant consequences of this for the way inverted indexes are built or used.
3. Consider the Dublin Core and a thesaurus (for example thesauri like the AAT (Art and Architecture Thesaurus), the ERIC thesaurus (Educational Resources Information Center), or MeSH (Medical Subject Headings). Briefly describe how they (DC and a thesaurus) differ from each other, and then describe how each can be used in information retrieval systems. Be sure to state which thesaurus you are describing.
 - (a) How they differ.
 - (b) How Dublin Core might be used in a search system.
 - (c) How a thesaurus might be used in a search system.
4. The article--“A Case For Interaction: A Study of Interactive Information Retrieval Behavior and Effectiveness” (Koenemann and Belkin)--distinguishes between opaque, transparent and penetrable feedback. Since even inexperienced users (in 1996) had notably better success using penetrable feedback, why isn't this approach seen more often in online search/retrieval tools?