

COMPUTER SCIENCE
COLLOQUIUM

Evaluating E-Discovery Search

Douglas W. Oard
University of Maryland

Abstract: Civil litigation in this country relies on each side making relevant evidence available to the other, a process known as "discovery." The explosive growth of information in digital form has led to an increasing focus on how search technology can best be applied to balance costs and responsiveness in what has come to be known as "e-discovery". This is now a multi-billion dollar business, one in which new vendors are entering the market frequently, usually with impressive claims about the efficacy of their products or services. Courts, attorneys, and companies are actively looking to understand what should constitute best practice, both in the design of search technology and in how that technology is employed. In this talk I will provide an overview of the e-discovery process, and then I will use that background to motivate a discussion of which aspects of that process the TREC Legal Track is seeking to model. I will then spend most of the talk describing two novel aspects of evaluation design: (1) recall-focused evaluation in large collections, and (2) modeling an interactive process for "responsive review" with fairly high fidelity. Although I will draw on the results of participating teams to illustrate what we have learned, my principal focus will be on discussing what we presently understand to be the strengths and weaknesses of our evaluation designs.

Bio:

Douglas Oard is a Professor at the University of Maryland, College Park, with joint appointments in the College of Information Studies and the Institute for Advanced Computer Studies, where he is the director of the Computational Linguistics and Information Processing Lab. Dr. Oard earned his Ph.D. in Electrical Engineering from the University of Maryland, and his research interests center around the use of emerging technologies to support information seeking by end users. His recent work has focused on interactive techniques for cross-language information retrieval, searching conversational media such as speech and email, evaluation design for e-discovery in the TREC Legal Track, and support for sense-making in large digital archival collections. Additional information is available at <http://terpconnect.umd.edu/oard/>.

Tuesday, September 20, 2011, 10:00 am
White Hall 208

MATHEMATICS AND COMPUTER SCIENCE
EMORY UNIVERSITY